

# Synthesizing Evidence: Data-Pooling as a Tool for Treatment Selection in Online Experiments

Zhenkang Peng

The Chinese University of Hong Kong, zhenkang.peng@cuhk.edu.hk

Chengzhang Li

Shanghai Jiao Tong University, cz.li@sjtu.edu.cn

Ying Rong

Shanghai Jiao Tong University, yrong@sjtu.edu.cn

Renyu (Philip) Zhang

The Chinese University of Hong Kong, philipzhang@cuhk.edu.hk

Randomized experiments are the gold standard for causal inference but face significant challenges in business applications, including limited traffic allocation, the need for heterogeneous treatment effect estimation, and the complexity of managing overlapping experiments. These factors lead to high variability in treatment effect estimates, making data-driven policy roll-out difficult. To address these issues, we introduce the data-pooling treatment roll-out (DPTR) framework, which enhances policy roll-out by pooling data across experiments rather than focusing narrowly on individual ones. DPTR can effectively accommodate both overlapping and non-overlapping traffic scenarios, regardless of linear or nonlinear model specifications. We demonstrate the framework's robustness through a three-pronged validation: (a) theoretical analysis shows that DPTR surpasses the traditional difference-in-mean and ordinary least squares methods under non-overlapping experiments, particularly when the number of experiments is large; (b) synthetic simulations confirm its adaptability in complex scenarios with overlapping traffic, rich covariates and nonlinear specifications; and (c) empirical applications to two experimental datasets from real-world platforms, demonstrating its effectiveness in guiding customized policy roll-outs for subgroups within a single experiment, as well as in coordinating policy deployments across multiple experiments with overlapping scenarios. By reducing estimation variability to improve decision-making effectiveness, DPTR provides a scalable, practical solution for online platforms to better leverage their experimental data in today's increasingly complex business environments.

*Key words:* Randomized Experiments, Data Pooling, Roll-out Policies, Experimentation on Online Platforms, Decision-aware Estimation.

---

## 1. Introduction

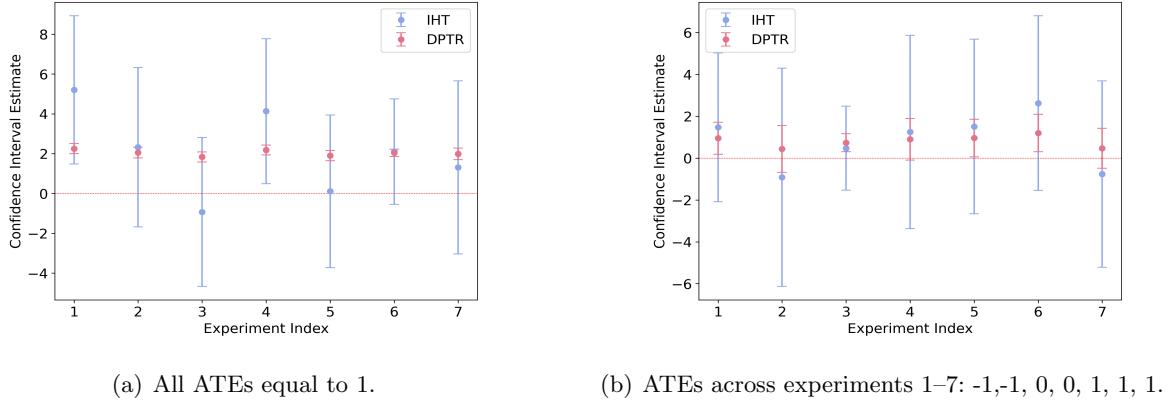
Randomized experiments have long been regarded as the gold standard for estimating causal effects across a wide range of scientific disciplines. The adoption of randomized experiments in business, especially in the tech sector, has gained significant momentum in recent years. Online platforms routinely use randomized control trials (RCTs) to shape a wide array of decisions, including the

product design, UI, recommendation algorithms, Ad placement, and pricing strategies (Luca and Bazerman 2021). The need for rapid validation and deployment has led companies to run hundreds or even thousands of RCTs concurrently. At Bing, for example, the number of completed experiments increased from fewer than 50 per week in 2008 to more than 300 per week by 2014 (Kohavi and Thomke 2017). Booking runs in excess of 1,000 concurrent experiments at any given moment across different products and target groups (Booking 2019). The widespread implementation of RCTs is not incidental: Kohavi et al. (2020) report that companies such as Microsoft, Google, and LinkedIn now conduct over 20,000 experiments annually.

A central issue faced by randomized experiments is data scarcity. Despite the large overall user base of many platforms, the amount of experimental traffic that can be allocated to any single treatment is often constrained, either due to operational limitations or risk concerns. As a result, the effective sample size for each experiment may be quite limited. For instance, Lewis and Rao (2015) examine 25 digital advertising RCTs conducted by large retailers and financial service firms. Their findings show that the median confidence interval for return on investment (ROI) is over 100 percentage points wide, making it nearly impossible for advertisers to distinguish between campaigns with a 50% ROI difference. This problem becomes more acute in the context of targeted experimentation and personalization, which are now standard practices in digital marketing. When experiments are stratified by user attributes to tailor interventions to specific segments, the resulting sample sizes for each subgroup can become vanishingly small, undermining the statistical power of the experiment (e.g., Athey and Imbens 2016, Lada et al. 2019). Furthermore, orthogonal experimental designs, often used to enable efficient estimation across high-dimensional treatment spaces, can exacerbate the issue by creating treatment combinations that are either underrepresented or completely unobserved in practice (see, Ye et al. 2025). These factors together pose a major barrier to identifying effective policy interventions and learning from past experiments, especially in fast-paced environments where decisions must be made with limited data and high uncertainty.

Given these challenges, our central research question is: *when running a large number of experiments and observing limited data for each, how should we improve the experiment roll-out decisions under data scarcity?* We propose a novel data-pooling treatment roll-out framework (DPTR). Instead of analyzing each experiment in isolation, DPTR integrates data across multiple experiments to enhance experiment roll-out decisions, where the “roll-out decision” refers to the policy selection in this paper, aiming at reducing variance with a tolerable increase in bias. In other words, this approach determines whether a policy should be implemented by leveraging both the data collected from its own experiment and the pooled data from other experiments. Specifically, the treatment effect estimation is conducted by combining an individual estimator and an anchor

estimator, defined as the average of all individual estimators, via a data-driven scale parameter. More importantly, the proposed DPTR framework is flexible and can be adapted to handle various experimentation scenarios, which differ along two dimensions: whether the experiments involve overlapping traffic and whether the underlying model is nonlinear. Beyond estimating average treatment effects (ATEs) by pooling data across experiments, the DPTR framework, by design, is also capable of accounting for heterogeneous treatment effects (HTEs) within subgroups in a single experiment. This enables DPTR framework to be broadly applicable to real-world business contexts, where experimentation scenarios are increasingly complex and data availability is limited.



**Figure 1 An simple numerical case study with seven experiments.**

Notes. Each experiment includes 10 observations, with 5 assigned to the control group and 5 to the treatment group. The noise term has a standard deviation of 3. We use the Difference-in-Means method for hypothesis testing.

**A case study.** We illustrate the advantage of DPTR framework over the conventional individual hypothesis testing (IHT) through a running example where seven experiments are conducted simultaneously. Each experiment collects 10 observations, and we compute confidence intervals of ATEs using both IHT and DPTR methods. We begin with the case with homogeneous ATEs, that is, all ATEs equal to 1. As shown in Figure 1 (a), the limited sample size causes IHT to produce excessively wide confidence intervals, with five lower bounds falling below zero. This implies that such a method fails to effectively identify experiments with positive ATEs. In contrast, our DPTR method significantly narrows the confidence intervals and enables effective identification of each experiment. Next, we consider a heterogeneous setting where ATEs differ across experiments, including both positive and negative values. As shown in Figure 1 (b), although our method may introduce some bias and occasionally misclassify non-positive ATEs as positive, the roll-out decisions guided by our method result in an improvement in overall reward compared to those based

on IHT. In particular, compared to IHT, DPTR successfully rolls out the 5th and 6th experiments with positive ATEs, although it fails to rule out the 1st experiment with a negative ATE.

The Bayesian framework offers a natural and principled method for aggregating evidence, and thus, similar to the proposed DPTR framework, can be employed to support data pooling across experiments to improve estimation efficiency at the level of individual experiments or targeted subgroups. Despite this alignment, the two frameworks differ fundamentally in their conceptual basis. The proposed DPTR framework, from a frequentist perspective, treats ATEs as fixed but unknown values, in contrast to the Bayesian approach that models ATEs as random variables. Moreover, the DPTR framework is more flexible in accommodating various data availability and model specifications: overlapping and non-overlapping experiment settings, as well as linear and nonlinear models. More importantly, the DPTR framework is decision-aware. That is, estimation and inference are directly aligned with downstream decisions, such as treatment roll-out decisions, by incorporating thresholds or significance levels into the analysis. This decision-centric design can lead to substantial performance gains in real-world deployment scenarios, where the goal is not merely to estimate effects accurately but to identify and implement effective interventions.

We demonstrate the value of the proposed DPTR framework from three perspectives. First, we theoretically examine DPTR's performance in the case with non-overlapping experiments and linear model specifications. We show that DPTR outperforms the conventional approach, which are based on IHT, when the number of experiments is sufficiently large. Second, we evaluate DPTR's performance using synthetic data across a variety of complex scenarios, including those involving nonlinear model specifications and overlapping experiments. Finally, we apply DPTR to two experimental datasets from real-world platforms, demonstrating its effectiveness in guiding customized policy roll-outs for subgroups within a single experiment, as well as in coordinating policy deployments across multiple experiments under overlapping scenarios.

Our contributions are threefold:

- 1. A general data-pooling treatment roll-out framework.** We propose a novel framework that enhances treatment selection by integrating data across multiple experiments. The treatment roll-out decision leverages both the data collected from its own experiment and the pooled data from other experiments. More importantly, the DPTR framework is decision-aware. That is, estimation and inference are directly aligned with downstream decisions, such as treatment roll-out decisions, by incorporating significance levels into the analysis. This framework is also broadly applicable, as it can handle overlapping and non-overlapping experiments, linear and nonlinear models, and supports personalized roll-out decisions. This general and flexible design makes DPTR well-suited for modern online platforms facing data scarcity and high-dimensional experimentation.

**2. Analysis and performance guaranteee.** We provide theoretical guarantees that DPTR outperforms IHT, particularly as the number of experiments increases. Specifically, we find that the average rewards associated with roll-out decisions derived by DPTR method with the optimal scale parameter are strictly higher than those of the IHT method. This implies that DPTR strikes a delicate balance between bias and variance by deriving the optimal scale parameter for shrinkage. We further construct a consistent estimator of the optimal scale parameter, which guarantees the superior performance of the proposed approach as the number of experiments increases. This ensures the robust performance of the proposed approach under data-driven settings. These analyses establish DPTR as a theoretically grounded method for improving the roll-out of experiments.

**3. Empirical validation and case studies with real-world data.** Through extensive numerical studies using synthetic data, we demonstrate the practical values of DPTR. Across diverse scenarios, DPTR consistently yields higher rewards and better decision quality than IHT and Bayesian benchmarks. In particular, the benefit of DPTR becomes more pronounced when the expected ATE is large, the number of experiments is large, and the sample size is small. In scenarios with covariate information and non-overlapping experiments, the personalized estimators prescribed by DPTR can yield additional improvements. Furthermore, we demonstrate the robustness of the DPTR method under model misspecification. Finally, the implementation of DPTR method to two datasets from real-world platforms shows that DPTR enables effective subgroup targeting and coordination across concurrent experiments, confirming its scalability and impact in real-world experimentation.

The rest of the paper is organized as follows. In Section 2, we review the related literature. In Section 3, we present our data-pooling treatment roll-out framework and prescribe implementation details of DPTR in four experimentation scenarios. In Section 4, we theoretically validate the proposed DPTR in the case with non-overlapping experiments under linear model specifications. In Section 5, we conduct comprehensive synthetic experiments to demonstrate the superior and robust performance of our proposed framework. In Section 6, we implement the framework to analyze real-world experiments. Section 7 concludes the paper.

## 2. Literature Review

Our work is related to two streams of literature: causal inference and its applications on online platforms, and small-data prediction and decision-making.

### 2.1. Causal inference and its application on online platforms

Randomized experiments have long been considered as the gold standard for estimating causal effects in social science research (e.g., Angrist and Pischke 2009). However, their implementation can be costly, and firms may sometimes face challenges in obtaining large sample sizes. To improve the efficiency of treatment effect estimation and reduce mean squared error (MSE), Rosenman

et al. (2023) and Gui (2024) propose a weighted sum that combines estimates from randomized experiments with secondary data. In addition, integrating experimental and secondary data offers additional advantages, such as aiding identification for estimating long-term effects (Athey et al. 2025, Imbens et al. 2025). While this line of research focuses on leveraging diverse data sources tied to a single policy, our approach aggregates data across multiple experimental policies.

Another approach to addressing the challenge of small sample sizes is Bayesian inference. The Bayesian approach offers several advantages, including its ability to handle uncertainty, incorporate prior knowledge, and model complex data structures. Recently, Bayesian approaches for causal inference have attracted growing attention (e.g., Imbens and Rubin 2015, Hahn et al. 2020). A number of studies have applied Bayesian methods to guide roll-out decisions (e.g., Abadie et al. 2023, Simester et al. 2025, Tetenov 2016). To ensure comparability with hypothesis testing from the frequentist perspective, Raftery (1995) proposed that a policy should be implemented in the Bayesian setting only when the posterior evidence is sufficiently strong, analogous to meeting a predefined significance level in the frequentist approach. The key distinction between our DPTR method and the Bayesian approach lies in the implementation of shrinkage: our method adopts a frequentist perspective, and explicitly incorporates the significance level into the shrinkage procedure applied to an unbiased estimator.

Our work is also related to the stream of studies on inference with multiple experiments. Conventional approaches for analyzing multiple experiments rely on factorial designs (e.g., Box et al. 1978, Wu and Hamada 2011). Recent works propose the potential outcome framework to facilitate causal inference across multiple experiments (Dasgupta et al. 2015, Pashley and Bind 2023). However, as pointed out by Ye et al. (2025), factorial designs become impractical in modern large-scale A/B testing environments, where the number of experiments is hundreds or thousands. Even if one adopts the fractional factorial design, only a limited number of treatment combinations are testable. To address this challenge, Ye et al. (2025) proposes a double machine learning framework that can infer all  $2^m$  treatment combinations using  $m + 2$  observed combinations. However, even when the ATEs are linear, the growing number of experiments on online platforms often leads to insufficient traffic per experiment. To address this, our paper focuses on how to make roll-out decisions with a large number of experiments and limited observations per experiment.

Our paper also contributes to the applications of causal inference to online platforms. The recent decade has witnessed a growing body of research on this topic. From an empirical perspective, field experiments on large-scale online platforms enable causal inference to empower decision making in a wide variety of business settings (e.g., Cheung et al. 2017, Cui et al. 2020, Zeng et al. 2023, Zhan et al. 2024). On the theoretical side, scholars develop novel methods to overcome challenges arising from experimentation and causal inference on online platforms, such as two-sided randomization

(e.g., Nandy et al. 2021, Johari et al. 2022, Ye et al. 2022), sequential experiments (e.g., Song and Sun 2024, Bojinov et al. 2023, Xiong et al. 2023, Ni et al. 2023, Ni 2025), block randomization (Candogan et al. 2021), multiple experiments (Ye et al. 2025), and personalized policy learning (Zhang et al. 2025). We contribute to this literature by proposing a new method to effectively pool data from multiple experiments and improve experiment roll-out decisions for online platforms.

## 2.2. Small-data prediction and decision-making

The pioneering work by Stein (1956) introduces the idea of data pooling for the simultaneous estimation of multiple Gaussian means and demonstrates its benefit over the decoupled approach, a result known as Stein’s phenomenon. This finding has spurred extensive follow-up research aimed at explaining and contextualizing Stein’s result (e.g., Brown 1971, Efron and Morris 1977). Building on this foundation, Gupta and Kallus (2022) extend Stein’s method to data-driven optimization problems, proposing a shrinkage-based approach that improves upon the decoupled approach by shrinking individual-level data to an anchor distribution. Lei et al. (2024) proposes to treat the aggregated top-level sales information as a regularization for fitting the individual-level prediction model, which improves forecasting performance. Chen et al. (2024) empirically investigate how data aggregation and sharing via a digital platform can enhance the analytics based on individual-level data for small retailers. A critical distinction lies in the nature of the data used: while problems like the newsvendor problem rely on observable labels such as random demand and cost parameters as direct inputs for optimization, our framework addresses situations where individual treatment effects are inherently unobservable. The lack of labeled outcomes calls for a tailored methodological approach to address the unique challenges of causal estimation and policy decision-making.

Our work is also related to multitask learning, which aims to learn both shared and task-specific representations across different tasks (Caruana 1997). In a similar vein, our work leverages observations across multiple experiments to improve estimation accuracy. While multi-task learning is primarily designed for predictive tasks, our method focuses on causal inference and is tailored towards decision-rule optimization. A related paradigm is transfer learning, a special case of multi-task learning, that improves a learner from one target domain by transferring information from a related source domain. Recently, transfer learning has been adopted to enhance the efficiency of operational decision-making (Bastani 2021, Nabi et al. 2022, Feng et al. 2023). The transfer learning approach relies on sufficient information from the related source domain to enhance predictions or decisions in the target domain. In contrast, our work involves multiple experiments, each with limited observations, and aims to improve the roll-out decisions for all experiments.

### 3. General Framework of Data Pooing in Experiment Roll Out

We now develop a new framework leveraging data pooling to improve the effectiveness of experiment roll-out decisions for online platforms. Suppose that a platform runs  $K$  independent A/B tests concurrently, each designed to evaluate the ATE of a distinct policy<sup>1</sup>. Without loss of generality, we assume each experiment is fully randomized, with users assigned to treatment and control conditions with equal probability. With multiple experiments, the interaction across overlapping treatments may exist. However, empirical evidence often finds that the interactions are relatively rare (Kohavi et al. 2013, Chan 2021, Microsoft 2023). Even if such interaction exists, the magnitude is relatively small, and thus, it does not overrule the main effect and affect the roll-out decisions (Chan 2021). Following the common practice of online platforms to balance speed and accuracy (CXL 2020), we assume that the policy ATEs are linearly additive<sup>2</sup>. Furthermore, we assume that the standard stable unit treatment value assumption (SUTVA) holds.

The goal of the platform is to identify and roll out *all* policies with a positive ATE. Let  $Y \in \mathbb{R}$  denote a key outcome variable the platform cares about (e.g., whether the user clicks the recommended advertisement), and  $D_k \in \{0, 1\}$  denote the treatment assignment of experiment  $k$ , capturing whether *all* the users are under the treatment or control condition. Define  $\mathbf{X} \in \mathbb{R}^{d_x}$  as the  $d_x$ -dimensional covariate vector of a user. Since the treatment effects of different policies are linearly separable, the ATE of policy  $k$  can be defined as:

$$\tau_k = \mathbb{E}[Y|D_k = 1] - \mathbb{E}[Y|D_k = 0],$$

where the expectation is taken with respect to  $(Y, \mathbf{X})$ . If the platform knows the ground-truth ATEs,  $\tau_1, \tau_2, \dots, \tau_K$ , it will roll out policy  $k$  if and only if  $\tau_k > 0$ . Hence, the optimal per-experiment reward of the platform is:

$$r^* = \frac{1}{K} \sum_{\tau_k > 0} \tau_k.$$

In practice, the ground-truth ATEs,  $\tau_1, \dots, \tau_K$ , are unobservable to the platform, so it runs A/B tests to estimate them and make roll-out decisions accordingly. Let  $\mathcal{S}$  denote the resulting experimental dataset, whose generation depends on the problem setting and experimentation method. For instance, when the  $K$  experiments are conducted independently, each user is assigned to either the treatment or the control condition of one experiment. In this case, if each experiment has  $N$  observations, the dataset can be represented as  $\mathcal{S} = \{(Y_{k,i}, D_{k,i}, X_{k,i}) : 1 \leq k \leq K, 1 \leq i \leq N\}$  where

<sup>1</sup> Alternatively, one could consider a single experiment designed to estimate HTEs across  $K$  different subgroups, aiming to determine whether the policy should be rolled out for each subgroup. Throughout this paper, except in Section 6.1, we adopt the notation of  $K$  experiments to illustrate DPTR framework. In Section 6.1, we demonstrate how this framework can be applied to roll out decisions across different subgroups using data from a single experiment.

<sup>2</sup> We also consider the case where the linear additivity does not hold to test the robustness of our framework in Section 5.5 and Section 6.2.

$Y_{k,i}$ ,  $D_{k,i}$  and  $X_{k,i}$  represent the individual outcome, treatment assignment status and covariate vector for subject  $i$  in experiment  $k$ , respectively. As another example, the platform may adopt an orthogonal experiment design in which a user may be simultaneously targeted by multiple experiments (e.g., Tang et al. 2010, Xiong et al. 2020). With a total of  $N$  users in the experiments, the dataset can be represented as  $\mathcal{S} = \{(Y_i, \mathbf{D}_i, X_i) : 1 \leq i \leq N\}$ , where  $\mathbf{D}_i = (D_{1,i}, D_{2,i}, \dots, D_{K,i})$  is the vector of treatment assignments across all  $K$  experiments for user  $i$ .  $Y_i$  and  $X_i$  denote the observed outcome and covariate vector for subject  $i$  under the joint realization of treatment assignments from all experiments.

Given the dataset  $\mathcal{S}$ , the platform typically relies on classical hypothesis testing to decide whether or not to roll out each experiment  $k$ . Specifically, the commonly adopted approaches to estimate and infer the ATEs of the policies include, e.g., difference-in-mean (DM) (e.g., Section 1.1 in Wager 2024), ordinary least squares (OLS) (e.g., Section 1.2 in Wager 2024), double machine learning (DML) (e.g., Chernozhukov et al. 2018, Farrell et al. 2020, Ye et al. 2025), etc. We provide the general procedure of such a standard decision-making framework for online platforms in Algorithm 1. Without loss of generality, we denote  $\mathcal{M}(\cdot)$  as a general hypothesis testing method for the null hypothesis  $H_0: \tau_k = 0$ , which maps the experimental dataset  $\mathcal{S}$  and significance level  $\alpha$  to the point estimate  $\hat{\tau}_k$  and the  $(1-\alpha)$ -confidence interval  $[\hat{\tau}_k^{\text{lb}}, \hat{\tau}_k^{\text{ub}}]$ . For example, in the special case where  $\hat{\tau}_k \sim \mathcal{N}(\tau_k, \sigma_k^2)$  and  $\sigma_k^2$  is known, we have  $\hat{\tau}_k^{\text{lb}} = \hat{\tau}_k - z_{1-\alpha/2}\sigma_k$  and  $\hat{\tau}_k^{\text{ub}} = \hat{\tau}_k + z_{1-\alpha/2}\sigma_k$ , where  $z_{1-\alpha/2}$  is the  $(1-\frac{\alpha}{2})$ -fractile quantile  $z$ -score of a standard normal distribution.

---

**Algorithm 1** Individual Hypothesis Testing (IHT)

---

**Require:** Set of policies  $[K] = \{1, 2, \dots, K\}$ ; experimental dataset  $\mathcal{S}$ ; significance level  $\alpha$ ; the classical hypothesis testing method  $\mathcal{M}(\cdot)$ .

```

1:  $\hat{\mathcal{A}}_{\text{IHT}} \leftarrow \{\}$ ; //Initialize the roll-out decision as an empty set.
2: for  $k \in [K]$  do
3:   Run  $\mathcal{M}(\mathcal{S}, k, \alpha)$  to obtain the point estimate  $\hat{\tau}_k$  and the  $(1-\alpha)$ -confidence interval  $[\hat{\tau}_k^{\text{lb}}, \hat{\tau}_k^{\text{ub}}]$ .
4:   if  $\hat{\tau}_k^{\text{lb}} > 0$  then
5:      $\hat{\mathcal{A}}_{\text{IHT}} \leftarrow \hat{\mathcal{A}}_{\text{IHT}} \cup \{k\}$ 
6:   end if
7: end for
```

**Ensure:**  $\hat{\mathcal{A}}_{\text{IHT}}$ : Roll out policy  $k$  if and only if  $k \in \hat{\mathcal{A}}_{\text{IHT}}$ .

---

The experiment roll-out strategy based on the IHT method (Algorithm 1) generates a realized per-experiment reward, which is given by:

$$\hat{r}_{\text{IHT}} = \frac{1}{K} \sum_{k \in \hat{\mathcal{A}}_{\text{IHT}}} \tau_k.$$

The IHT will have great performance if the sample size of each experiment,  $N$ , is large. However, if  $N$  is small, the variance of ATE estimator  $\hat{\tau}_k$  of experiment  $k$ , will be too large, resulting in a poor performance of per-experiment reward for the IHT method. To address this challenge, we design a new estimator that combines data from different experiments to lower its variance, at the cost of a higher bias. Based on the idea of shrinkage (e.g., Gupta and Kallus 2022), the new estimator of  $\tau_k$  is parametrized by an anchor  $\tau$  and a scale parameter  $\beta \geq 0$ :

$$\bar{\tau}_k = \frac{N}{N+\beta} \hat{\tau}_k + \frac{\beta}{N+\beta} \tau. \quad (1)$$

Based on the new estimator (1), we devise the platform roll-out decision according to a new null hypothesis  $\bar{H}_0 : \frac{N}{N+\beta} \tau_k + \frac{\beta}{N+\beta} \tau = 0$  at the significance level  $\alpha$ . Therefore, the  $(1-\alpha)$ -confidence interval for the new point estimate  $\bar{\tau}_k$  is given by  $[\bar{\tau}_k^{\text{lb}}, \bar{\tau}_k^{\text{ub}}]$  where  $\bar{\tau}_k^{\text{lb}} = \frac{N}{N+\beta} \hat{\tau}_k^{\text{lb}} + \frac{\beta}{N+\beta} \tau$  and  $\bar{\tau}_k^{\text{ub}} = \frac{N}{N+\beta} \hat{\tau}_k^{\text{ub}} + \frac{\beta}{N+\beta} \tau$ . Here, we set  $\tau = \hat{\tau}_0 := \frac{1}{K} \sum_k \hat{\tau}_k$ , as this choice is a least squares estimator of central tendency among all individual estimators  $\hat{\tau}_k$ . Next, we are ready to propose a general framework to identify proper values for  $\beta$ , so as to optimize the experiment roll-out decisions, as detailed in Algorithm 2.

---

**Algorithm 2** Data-Pooling Treatment Roll-Outs (DPTR)

---

**Require:** Set of policies  $[K] = \{1, 2, \dots, K\}$ ; experimental dataset  $\mathcal{S}$ ; significance level  $\alpha$ ; the classical hypothesis testing method  $\mathcal{M}(\cdot)$ .

- 1:  $\hat{\mathcal{A}}_{\text{DPTR}} \leftarrow \{\}$ ; //Initialize the roll-out decision as an empty set.
- 2:  $\{(\hat{\tau}_k, \hat{\tau}_k^{\text{lb}}, \hat{\tau}_k^{\text{ub}}) : k \in [K]\} \leftarrow \{\mathcal{M}(\mathcal{S}, k, \alpha) : k \in [K]\}$ . //Implement the classical hypothesis testing
- 3: Obtain data-driven parameters  $\hat{\tau}_0 \leftarrow \frac{1}{K} \sum_k \hat{\tau}_k$  and  $\hat{\beta}(\mathcal{S}, \mathcal{M})$
- 4: **for**  $k \in [K]$  **do**
- 5:      $\bar{\tau}_k^{\text{lb}} \leftarrow \frac{N}{N+\hat{\beta}(\mathcal{S}, \mathcal{M})} \hat{\tau}_k^{\text{lb}} + \frac{\hat{\beta}(\mathcal{S}, \mathcal{M})}{N+\hat{\beta}(\mathcal{S}, \mathcal{M})} \hat{\tau}_0$ ,  $\bar{\tau}_k^{\text{ub}} \leftarrow \frac{N}{N+\hat{\beta}(\mathcal{S}, \mathcal{M})} \hat{\tau}_k^{\text{ub}} + \frac{\hat{\beta}(\mathcal{S}, \mathcal{M})}{N+\hat{\beta}(\mathcal{S}, \mathcal{M})} \hat{\tau}_0$ . //Construct the new lower bound and upper bound
- 6:     **if**  $\bar{\tau}_k^{\text{lb}} > 0$  **then**
- 7:          $\hat{\mathcal{A}}_{\text{DPTR}} \leftarrow \hat{\mathcal{A}}_{\text{DPTR}} \cup \{k\}$
- 8:     **end if**
- 9: **end for**

**Ensure:**  $\hat{\mathcal{A}}_{\text{DPTR}}$ : Roll out policy  $k$  if and only if  $k \in \hat{\mathcal{A}}_{\text{DPTR}}$ .

---

Algorithm 2 provides a general hypothesis testing procedure with data pooling to roll out experiments. In Algorithm 2, when constructing the confidence interval for the new estimator  $\bar{\tau}_k$ , we ignore the randomness of  $\hat{\tau}_0$  and  $\hat{\beta}(\mathcal{S}, \mathcal{M})$ , which are obtained from the data of all  $K$  experiments. When  $K$  is large, the variances of  $\hat{\tau}_0$  and  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  are orders of magnitude smaller than that of

$\hat{\tau}_k$ . Moreover, Algorithm 2 does not specify the formula of  $\hat{\beta}(\mathcal{S}, \mathcal{M})$ , which depends on the specific context, the dataset  $\mathcal{S}$ , and the hypothesis testing method  $\mathcal{M}$ . Intuitively,  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  is larger when (1) The variation in individual treatment effects within each experiment is large, i.e., the individual ATE estimates ( $\hat{\tau}_k$ 's) are more volatile and less credible; (2) The treatment effects of different experiments ( $\tau_k$ 's) are concentrated, so that the data-driven anchor  $\hat{\tau}_0$  effectively aggregates information across different experiments, thus significantly enhancing the reliability of  $\bar{\tau}_k$ . This intuition is formally derived in Theorem 1 and Theorem 2, where we consider the simplest case with no overlapping traffic, no covariate information and linear model specifications. Furthermore,  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  may vary across experiments due to differences in experiment-specific covariate information, assuming there is no overlapping traffic. This intuition is formally established in Theorem 4 and Theorem 5. In addition, throughout Sections 3.1 to 3.4, we present the corresponding formulas of  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  for experiments with overlapping and nonoverlapping subjects as well as linear and nonlinear model specifications. Given  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  and  $\hat{\tau}_0$ , the experiment roll-out strategy based on the DPTR method (Algorithm 2) generates a realized per-experiment reward:

$$\hat{r}_{\text{DPTR}} = \frac{1}{K} \sum_{k \in \hat{\mathcal{A}}_{\text{DPTR}}} \tau_k.$$

### 3.1. Scenario 1: Non-Overlapping Experiments With Linear Specifications

We begin by examining a scenario where  $K$  experiments are conducted in  $K$  separate subject pools, with each pool exclusively assigned to a single experiment. In this scenario, we consider linear model specifications. Specifically, we assume the following data-generating process (DGP):

$$Y_{k,i} = a_k + \tau_k D_{k,i} + \epsilon_{k,i}, \quad k = 1, 2, \dots, K, \quad i = 1, \dots, N, \quad (2)$$

where  $\tau_k$  represents the ATE of policy  $k$ ,  $\epsilon_{k,i}$  is the i.i.d. random noise with zero mean and variance  $\sigma_k^2$ , and  $a_k$  denotes the expected outcome under control condition for experiment  $k$ .

For each experiment, suppose the platform allocates exactly  $N$  users exclusively to it, with  $N/2$  randomly assigned to the treatment condition and  $N/2$  to the control condition. The total dataset in this scenario can be represented as  $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K\}$ , where  $\mathcal{S}_k = \{(Y_{k,i}, D_{k,i}) : 1 \leq i \leq N\}$ . In this case, the classic hypothesis testing method  $\mathcal{M}(\cdot)$  can be the unbiased DM estimator:

$$\hat{\tau}_k := \frac{2}{N} \sum_{D_{k,i}=1} Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=0} Y_{k,i}. \quad (3)$$

It is straightforward to derive that  $\text{Var}(\hat{\tau}_k) = \frac{4\sigma_k^2}{N}$ . Furthermore, the unbiased estimator for variance  $\sigma_k^2$  in experiment  $k$  can be expressed as:

$$s_k^2 = \frac{1}{N-2} \sum_{j \in \{0,1\}} \sum_{D_{k,i}=j} (Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=j} Y_{k,i})^2.$$

Thus, for any experiment  $k$ , we have the central limit theorem (CLT):

$$\sqrt{N}(4s_k^2)^{-1/2}(\hat{\tau}_k - \tau_k) \rightarrow_d \mathcal{N}(0, 1),$$

where  $\rightarrow_d$  refers to convergence in distribution. Based on the estimators  $\{\hat{\tau}_1, \dots, \hat{\tau}_K\}$  and  $\{4s_1^2, \dots, 4s_K^2\}$ , we construct  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  as follows:

$$\hat{\beta} = \frac{\frac{1}{K} \sum_k 4s_k^2}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k 4s_k^2} + \frac{z_{1-\alpha/2} \sqrt{N \frac{1}{K} \sum_k 4s_k^2}}{\hat{\tau}_0}. \quad (4)$$

In the first term of  $\hat{\beta}$ , the numerator  $\frac{1}{K} \sum_k 4s_k^2$  is an unbiased estimator of  $\frac{1}{K} \sum_k 4\sigma_k^2$ , capturing the average variation of all individual estimators  $\hat{\tau}_k$ . Hence, more variable estimations of the treatment effects for individual experiments lead to a larger  $\hat{\beta}$ , which in turn shrinks the new estimator  $\bar{\tau}_k$  (recall Eqn. (1)) more towards  $\hat{\tau}_0$ . The denominator  $\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k 4s_k^2$  is an unbiased estimator for  $\frac{1}{K} \sum_k (\tau_k - \frac{1}{K} \sum_k \tau_k)^2$  (see Theorem 2), capturing the variability of different experiments' ATEs. When the ATEs across different experiments are more concentrated, the aggregated information provided by  $\hat{\tau}_0$  becomes more valuable, so the scale parameter  $\hat{\beta}$  is larger and  $\bar{\tau}_k$  is shrunk to  $\hat{\tau}_0$  further. The second term of  $\hat{\beta}$  functions as a necessary adjustment for hypothesis testing. We relegate the derivation for  $\hat{\beta}$ 's formula as Eqn. (4) to Theorem 1 (see Section 4.1).

Next, we incorporate covariate information into the OLS model specification and DGP as follows:

$$Y_{k,i} = a_k + \tau_k D_{k,i} + \boldsymbol{\theta}_k^\top X_{k,i} + \epsilon_{k,i}, \quad 1 \leq k \leq K, 1 \leq i \leq N. \quad (5)$$

We denote  $\mathbf{I} = [0, 1, \dots, 0] \in \mathbb{R}^{2+d_x}$ ,  $\mathbf{Y}_k = [Y_{k,1}, \dots, Y_{k,N}]^\top$ ,  $\boldsymbol{\epsilon}_k = [\epsilon_{k,1}, \dots, \epsilon_{k,N}]^\top$  and  $\mathbf{t}_k = [(1, D_{k,1}, \mathbf{X}_{k,1}), \dots, (1, D_{k,N}, \mathbf{X}_{k,N})]^\top$ . Thus, we can obtain the estimator  $\hat{\tau}_k$  by OLS as follows:

$$\hat{\tau}_k = \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \mathbf{Y}_k = \tau_k + \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \boldsymbol{\epsilon}_k. \quad (6)$$

It is straightforward to derive that  $\text{Var}(\hat{\tau}_k) = \sigma_k^2 \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{I}$ . Furthermore, the unbiased estimator for variance  $\sigma_k^2$  in experiment  $k$  can be expressed as:

$$s_k^2 = \frac{1}{N - 2 - d_x} (\mathbf{Y}_k - \mathbf{t}_k (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \mathbf{Y}_k)^\top (\mathbf{Y}_k - \mathbf{t}_k (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \mathbf{Y}_k).$$

For any experiment  $k$ , we have the following CLT:

$$\sqrt{N}(b_k^2 s_k^2)^{-1/2}(\hat{\tau}_k - \tau_k) \rightarrow_d \mathcal{N}(0, 1), \quad (7)$$

where  $b_k = \sqrt{N \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{I}}$ . Similarly, according to the Eqn. (4), one can construct the same  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  for all experiments as follows:

$$\hat{\beta} = \frac{\frac{1}{K} \sum_k b_k^2 s_k^2}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k b_k^2 s_k^2} + \frac{z_{1-\alpha/2} \sqrt{N \frac{1}{K} \sum_k b_k^2 s_k^2}}{\hat{\tau}_0}. \quad (8)$$

Furthermore, we can observe that  $\{b_k^2 : k \in [K]\}$  are different across experiments and they can be derived from the training covariate vectors, which are known prior to making the roll-out decision. Thus, we can derive a personalized  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  for experiment  $k$  as follows:

$$\hat{\beta}_k = \frac{b_k^2 \left( \frac{1}{K} \sum_{k'} s_{k'}^2 \right)}{\frac{1}{K} \sum_{k'} (\hat{\tau}_{k'} - \hat{\tau}_0)^2 - \frac{1}{N} \left( \frac{1}{K} \sum_{k'} b_{k'}^2 \cdot \frac{1}{K} \sum_{k'} s_{k'}^2 \right)} + \frac{z_{1-\alpha/2} b_k \sqrt{N \frac{1}{K} \sum_{k'} s_{k'}^2}}{\hat{\tau}_0}. \quad (9)$$

We defer the derivation of the formula for  $\hat{\beta}_k$ , presented in Eqn. (9), to Theorem 4 (see Section 4.2), and demonstrate that the personalized  $\hat{\beta}_k$  performs even better than the shared  $\hat{\beta}$  in Eqn. (8) using simulations with synthetic data (see Section 5.1).

### 3.2. Scenario 2: Non-Overlapping Experiments With Nonlinear Specifications

Building upon the scenario outlined in Section 3.1, we now introduce a second scenario that considers nonlinear model specifications, extending the framework to a partial linear model. Specifically, we assume the following DGP:

$$Y_{k,i} = g_k(X_{k,i})^\top \mathbf{t}_{k,i} + \epsilon_{k,i}, \quad k = 1, 2, \dots, K, \quad i = 1, \dots, N,$$

where  $g_k(\cdot) : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^2$  represents the true response function for experiment  $k$ , and  $\mathbf{t}_{k,i} = [1, D_{k,i}]^\top$  is the treatment vector, which includes the constant term. The term  $\epsilon_{k,i}$  denotes i.i.d. random noise with zero mean and variance  $\sigma_k^2$ . All functions  $\{g_1, g_2, \dots, g_K\}$  belong to the same function class  $\mathcal{F}$ . Consequently, the total dataset in this scenario is given by  $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K\}$ , where  $\mathcal{S}_k = \{(Y_{k,i}, D_{k,i}, X_{k,i}) : i = 1, \dots, N\}$ . The ATE for experiment  $k$  is denoted as  $\tau_k = \mathbb{E}[g_k(X_k)^\top \mathbf{t}^*]$ , where  $\mathbf{t}^* = [0, 1]^\top$ .

Building on the partial linear framework, we employ the double machine learning method (Farrell et al. 2020) as  $\mathcal{M}$  in this scenario. To begin, we define the loss function for estimation and inference function as:

$$\ell(Y, \mathbf{t}, f(X)) = (Y - f(X)^\top \mathbf{t})^2. \quad H(X, f(X) : \mathbf{t}^*) = f(X)^\top \mathbf{t}^*.$$

Here, for each experiment  $k$ , we apply the cross-fitting techniques (Chernozhukov et al. 2018, Farrell et al. 2020) to obtain the estimator  $\hat{\tau}_k$ . Specifically, we define,

$$\psi(Y, X, \mathbf{t}, f(\cdot), \Lambda) = H(X, f(X) : \mathbf{t}^*) - H_f(X, f(X) : \mathbf{t}^*) \Lambda(X)^{-1} \ell_f(Y, \mathbf{t}, f(X)), \quad (10)$$

where  $H_f$  and  $\ell_f$  are the gradients of  $H$  and  $\ell$  with respect to  $f$ , and  $\Lambda(X) = \mathbb{E}[\ell_{ff}(Y, \mathbf{t}, f(X)) | X]$  represents the conditional expectation of the Hessian of  $\ell$ . The expectation of  $H(X, f(X); \mathbf{t}^*)$  represents the true ATE we aim to estimate. However, in practice, due to the complexity and regularization of  $f(X)$ , the sample mean of  $H(X, f(X); \mathbf{t}^*)$  does not yield an unbiased estimator. Therefore, the second term in Eqn. (10) was introduced to correct the bias.

The estimation and experiment roll-out process can be summarized as follows. First, the data set  $\mathcal{S}_k$  is split into  $S$  subsets with equal size, denoted by  $\mathcal{S}_{k,s}$  where  $s \in \{1, \dots, S\}$ . Let  $\mathcal{S}_{k,s}^c$  be the complement of  $\mathcal{S}_{k,s}$ . Then, for each  $s \in \{1, \dots, S\}$ , we use  $\mathcal{S}_{k,s}^c$  to estimate  $\Lambda_k(X_k) = \mathbb{E}[\ell_{ff}(Y_k, \mathbf{t}_k, g_k(X_k))|X_k]$  and  $g_k(\cdot)$ . We denote  $\hat{g}_{k,s}(\cdot)$  and  $\hat{\Lambda}_{k,s}(\cdot)$  as the estimators for  $g_{k,s}(\cdot)$  and  $\Lambda_{k,s}(\cdot)$ , respectively. Then, based on Eqn. (10), the final estimator  $\hat{\tau}_k$  can be written as:

$$\hat{\tau}_k = \frac{1}{S} \sum_{s=1}^S \hat{\tau}_{k,s}, \quad \hat{\tau}_{k,s} = \frac{1}{|\mathcal{S}_{k,s}|} \sum_{i \in \mathcal{S}_{k,s}} \psi(Y_{k,i}, X_{k,i}, \mathbf{t}_{k,i}, \hat{g}_{k,s}(X_{k,i}), \hat{\Lambda}_{k,s}(X_{k,i})).$$

The estimator for variance of  $\hat{\tau}_k$  can be written as:

$$\hat{\Psi}_k = \frac{1}{S} \sum_{s=1}^S \hat{\Psi}_{k,s}, \quad \hat{\Psi}_{k,s} = \frac{1}{|\mathcal{S}_{k,s}|} \sum_{i \in \mathcal{S}_{k,s}} \left( \psi(Y_{k,i}, X_{k,i}, \mathbf{t}_{k,i}, \hat{g}_{k,s}(X_{k,i}), \hat{\Lambda}_{k,s}(X_{k,i})) - \hat{\tau}_k \right)^2.$$

Based on Theorem 3 of [Farrell et al. \(2020\)](#), as long as the nuisance parameter estimator  $\hat{g}_k(\cdot)$  converges to  $g_k(\cdot)$  sufficiently fast, we have:

$$\sqrt{N} \hat{\Psi}_k^{-1/2} (\hat{\tau}_k - \tau_k) \rightarrow_d \mathcal{N}(0, 1). \quad (11)$$

Based on the estimators  $\{\hat{\tau}_1, \dots, \hat{\tau}_K\}$  and  $\{\hat{\Psi}_1, \dots, \hat{\Psi}_K\}$ , we can construct the scale parameter  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  following the same intuition as Eqn. (8):

$$\hat{\beta} = \frac{\frac{1}{K} \sum_k \hat{\Psi}_k}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k \hat{\Psi}_k} + \frac{z_{1-\alpha/2} \sqrt{N \frac{1}{K} \sum_k \hat{\Psi}_k}}{\hat{\tau}_0}. \quad (12)$$

Similar to Eqn. (9), we can also construct the heuristic personalized scale parameter  $\hat{\beta}_k$ 's in this setting. First, we can compute  $b_k = \sqrt{N \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{I}}$  using the covariate information. Second, by comparing Eqn. (7) and (11), we can find that  $b_k^2 s_k^2$  and  $\hat{\Psi}_k$  play the same role and we can construct  $s_k^2 = \frac{\hat{\Psi}_k}{b_k^2}$ . Thus, we can derive a personalized  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  for experiment  $k$  as follows:

$$\hat{\beta}_k = \frac{b_k^2 \left( \frac{1}{K} \sum_{k'} \frac{\hat{\Psi}_{k'}}{b_{k'}^2} \right)}{\frac{1}{K} \sum_{k'} (\hat{\tau}_{k'} - \hat{\tau}_0)^2 - \frac{1}{N} \left( \frac{1}{K} \sum_{k'} b_{k'}^2 \cdot \frac{1}{K} \sum_{k'} \frac{\hat{\Psi}_{k'}}{b_{k'}^2} \right)} + \frac{z_{1-\alpha/2} b_k \sqrt{N \frac{1}{K} \sum_{k'} \frac{\hat{\Psi}_{k'}}{b_{k'}^2}}}{\hat{\tau}_0}. \quad (13)$$

### 3.3. Scenario 3: Overlapping Experiments With Linear Specifications

In this subsection, we examine the scenario that a user may be simultaneously targeted by multiple experiments. Similarly, we first examine the scenario with linear model specifications, followed by an extension that considers nonlinear model specifications in the next subsection. Specifically, we assume the following DGP:

$$Y_i = a + \sum_{k \in K} \tau_k D_{k,i} + \epsilon_i, \quad i = 1, 2, \dots, N, \quad (14)$$

where  $\tau_k$  represents the ATE of policy  $k$ ,  $\epsilon_i$  is the i.i.d. random noise with zero mean and variance  $\sigma^2$ , and  $a$  represents the expected outcome if a subject receives the control status in all experiments. Furthermore,  $D_{k,i}$ 's are i.i.d. Bernoulli random variables with  $\mathbb{P}[D_{k,i} = 1] = \mathbb{P}[D_{k,i} = 0] = 0.5$ .

In this scenario, the dataset can be represented as  $\mathcal{S} = \{(Y_i, \mathbf{D}_i) : i = 1, \dots, N\}$  where  $\mathbf{D}_i = [D_{1,i}, D_{2,i}, \dots, D_{K,i}]^\top$  denotes the treatment status vector across all experiments. We define  $\boldsymbol{\tau} = [a, \tau_1, \tau_2, \dots, \tau_K]^\top$  as the average treatment effect vector. The classical hypothesis testing method  $\mathcal{M}(\cdot)$  in this scenario adopts OLS to estimate  $\boldsymbol{\tau}$ .

To proceed, we define  $\mathbf{t}_i = [1, \mathbf{D}_i]^\top$ . Let  $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]^\top$  and  $\mathcal{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N]^\top$ . Thus, the OLS estimator of  $\boldsymbol{\tau}$  is given by:

$$\hat{\boldsymbol{\tau}} = (\mathcal{T}^\top \mathcal{T})^{-1} \mathcal{T}^\top \mathbf{Y}.$$

In addition, we have,  $\mathbb{E}(\hat{\boldsymbol{\tau}}) = \boldsymbol{\tau}$  and  $\mathbb{V}(\hat{\boldsymbol{\tau}}|\mathcal{T}) = \sigma^2(\mathcal{T}^\top \mathcal{T})^{-1}$ . To estimate the variance  $\sigma^2$ , we use  $\hat{\sigma}^2 = \frac{\sum_i (Y_i - \hat{\boldsymbol{\tau}}^\top \mathbf{t}_i)^2}{N-K-1}$ . Next, let  $I_k$  be the  $(K+1)$ -dimensional vector where the  $(k+1)$ th component is equal to 1, and all other components are equal to 0. The estimate of the average treatment effect for experiment  $k$  is given by,  $\hat{\tau}_k = I_k^\top \hat{\boldsymbol{\tau}}$ . The standard error (SE) of  $\hat{\tau}_k$  is,  $\text{SE}_k = \sqrt{I_k^\top \hat{\sigma}^2 (\mathcal{T}^\top \mathcal{T})^{-1} I_k}$ , and we denote  $s_k^2 = \frac{N}{4} \text{SE}_k^2$ . Similarly, by Greene (2003), we can conclude that:

$$\sqrt{N}(4s_k^2)^{-1/2}(\hat{\tau}_k - \tau_k) \xrightarrow{d} \mathcal{N}(0, 1).$$

Similar to Section 3.1, we can derive the scale parameter  $\hat{\beta}(\mathcal{S}, \mathcal{M})$  as:

$$\hat{\beta} = \frac{\frac{1}{K} \sum_k 4s_k^2}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k 4s_k^2} + \frac{z_{1-\alpha/2} \sqrt{N \frac{1}{K} \sum_k 4s_k^2}}{\hat{\tau}_0}. \quad (15)$$

When covariate vectors are incorporated in the OLS model with the following DGP:

$$Y_i = a + \sum_k \tau_k D_{k,i} + \boldsymbol{\theta}^\top X_i + \epsilon_i, \quad 1 \leq i \leq N, \quad (16)$$

the OLS estimator for each experiment  $k$ ,  $\hat{\tau}_k$ , is asymptotically normal estimator, so Algorithm 2 can be applied with the scale parameter defined by Eqn. (15).

### 3.4. Scenario 4: Overlapping Experiments With Nonlinear Specifications

Finally, we consider the setting with overlapping experiments and nonlinear model specifications. Similar to Section 3.2, we adopt the partial linear model framework, and the DGP is given by:

$$Y_i = \mathbf{g}(X_i)^\top \mathbf{t}_i + \epsilon_i, \quad i = 1, \dots, N, \quad (17)$$

where  $\mathbf{g}(\cdot) : \mathbb{R}^{dx} \rightarrow \mathbb{R}^{K+1}$  is the true response function and  $\mathbf{t}_i = [1, \mathbf{D}_i]^\top$  is the treatment vector including the intercept.  $\epsilon_i$  denotes the i.i.d. random noise with zero mean and variance  $\sigma^2$ . The

dataset in this scenario can be represented as  $\mathcal{S} = \{(Y_i, X_i, \mathbf{t}_i) : i = 1, \dots, N\}$ . The ATE for experiment  $k$  can be denoted as  $\tau_k = \mathbb{E}[\mathbf{g}(X)^\top \mathbf{t}_k^*]$  where  $\mathbf{t}_k^*$  is a  $K + 1$  dimension vector of which the  $(k + 1)$ th component is equal to 1 and other components are equal to zero.

The overlapping scenario does not affect the validity of Theorem 3 in [Farrell et al. \(2020\)](#). Similar to Section 3.2, we obtain the ATE estimator  $\hat{\tau}_k$  and the corresponding variance estimator  $\hat{\Psi}_k$ , along with the asymptotic normality result:

$$\sqrt{N}\hat{\Psi}_k^{-1/2}(\hat{\tau}_k - \tau_k) \rightarrow_d \mathcal{N}(0, 1).$$

Then, we can construct the scale parameter similar to Eqn. (12):

$$\hat{\beta}(\mathcal{S}) = \frac{\frac{1}{K} \sum_k \hat{\Psi}_k}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k \hat{\Psi}_k} + \frac{z_{1-\alpha/2} \sqrt{N \frac{1}{K} \sum_k \hat{\Psi}_k}}{\hat{\tau}_0}.$$

For the rest of this paper, we will demonstrate that the decisions derived from hypothesis testing based on the DPTR method in Algorithm 2 outperform those obtained using the traditional IHT method in Algorithm 1 across various scenarios, through a multi-facet analysis, including theoretical proofs, numerical experiments, and empirical applications.

## 4. Theoretical Analysis

In this section, we derive the optimal scale parameter  $\beta$  and provide the theoretical justification for the DPTR framework. To this end, we focus on the simplest scenario, non-overlapping experiments under linear model specifications (see Section 3.1 for details). Specifically, we first prove that, in the case without covariate information, i.e., the DGP follows Eqn. (2), the DPTR experiment roll-out method generates a higher reward than the IHT method. We then extend this result to the setting with covariates (see Eqn. (5)). The following assumption is made throughout this section.

**ASSUMPTION 1.**  $\tau_k$  is drawn from a normal distribution,  $\mathcal{N}(\tau_0, \sigma_0^2)$ . Furthermore, the i.i.d. random noise  $\epsilon_{k,i}$  follows a normal distribution,  $\mathcal{N}(0, \sigma^2)$ .

In practice, when the platform conducts a randomized field experiment to test a specific policy, it will typically do some pilot study to assess its potential value. In most cases, the platform will proceed with the field experiment only if the policy's ATE is likely to be positive. Therefore, in general, the expected ATE across all field experiments should be positive, i.e.,  $\tau_0 > 0$ .

### 4.1. Model Without Covariates

In this subsection, we theoretically justify our proposed method for the setting without covariates. In this setting, the DM estimator (see Eqn. (3)) follows a normal distribution with mean  $\tau_k$  and variance  $4\sigma^2/N$  under Assumption 1.

Suppose that the variance  $\sigma^2$  is known. The platform uses  $z$ -statistics to construct the confidence level in  $\mathcal{M}(\cdot)$ . We define the per-experiment reward of DPTR with scale  $\beta$  and anchor  $\tau$  as follows:

$$\tilde{r}(\beta, \tau) = \frac{1}{K} \sum_{\tilde{\tau}_k > \frac{N}{N+\beta} \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}}} \tau_k. \quad (18)$$

It follows that  $\tilde{r}(0, \cdot) = \tilde{r}(0, \tau_0) = \hat{r}_{IHT}$ . Define  $\mathcal{R}(\beta, \tau) := \lim_{K \uparrow +\infty} \mathbb{E}[\tilde{r}(\beta, \tau)]$  as the expected per experiment reward when the number of experiments  $K \rightarrow \infty$ . Our analysis begins with identifying the optimal value of  $\beta$  that maximizes the expected per-experiment reward when  $\tau = \tau_0$ .

**THEOREM 1.** *Suppose that Assumption 1 holds and  $\sigma$  is known. If  $\tau = \tau_0$ , the optimal value of  $\beta$  is:*

$$\beta^* := \arg \max_{\beta \geq 0} \mathcal{R}(\beta, \tau_0) = \frac{4\sigma^2}{\sigma_0^2} + \frac{2\sqrt{N}z_{1-\alpha/2}\sigma}{\tau_0}. \quad (19)$$

In particular,  $\mathcal{R}(\beta^*, \tau_0) > \mathcal{R}(0, \tau_0) = \lim_{K \uparrow +\infty} \mathbb{E}[\hat{r}_{IHT}]$ .

Theorem 1 characterizes the optimal scale parameter  $\beta^*$  that maximizes the expected per-experiment reward of the DPTR roll-out method when the anchor is set at  $\tau = \tau_0$ . As a consequence, our proposed DPTR method with the optimal scale parameter outperforms the classical DM method, highlighting the potential value of data pooling for multiple A/B tests.

We can treat  $\hat{\tau}_k$  as the signal from experiment  $k$ , and  $\tau_0$  as the aggregate information across all experiments. Based on Eqn. (1), we observe that, when  $N$  is fixed, a larger value of  $\beta$  results in less weight being assigned to the individual signal under the DPTR method. Eqn. (19) prescribes that  $\beta^*$  consists of two positive terms  $\frac{4\sigma^2}{\sigma_0^2}$  and  $\frac{2\sqrt{N}z_{1-\alpha/2}\sigma}{\tau_0}$ . The first term is proportional to the ratio of the variance of noise within an experiment to the variance of treatment effects across experiments. When this ratio is large, it indicates that the aggregate information from all experiments is more reliable than the individual signal. In this case, less weight should be placed on the individual signal. The second term may seem counterintuitive at the first glance: why does a smaller  $\tau_0$  imply less weight being placed on the individual signal? In fact, the weight on the individual signal depends on the relative ratio of  $\tau_0$  to  $\sigma$ . If the individual signals have a small variance, they may be more informative than the aggregate signal  $\tau_0$ . Therefore, when the individual signal is more precise, it remains more beneficial to assign greater weight to it, even when  $\tau_0$  is small. Although  $\beta^*$  is increasing in  $N$ , Eqn. (1) also implies that the weight placed on the individual signal also increases with  $N$  under DPTR.

While the parameters  $\tau_0$ ,  $\sigma^2$ , and  $\sigma_0^2$ , which are used for deriving the optimal scale parameter, are unobservable in practice, one can leverage the pooled data from all experiments to estimate them. Thus, a natural estimator of the optimal scale parameter is one that replaces these parameters in Eqn. (19) by their estimates, as formally stated in Theorem 2.

**THEOREM 2.** Suppose Assumption 1 holds. Define  $s_k^2 := \frac{1}{N-2} \sum_{j \in \{0,1\}} \sum_{D_{k,i}=j} (Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=j} Y_{k,i})^2$  and

$$\hat{\beta}^* := \frac{\frac{1}{K} \sum_k 4s_k^2}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k 4s_k^2} + \frac{z_{1-\alpha/2} \sqrt{N \frac{1}{K} \sum_k 4s_k^2}}{\hat{\tau}_0}. \quad (20)$$

We have, as  $K \uparrow \infty$ ,  $\hat{\tau}_0 \rightarrow_p \tau_0$  and  $\hat{\beta}^* \rightarrow_p \beta^*$ , where  $\rightarrow_p$  refers to convergence in probability.

As shown in Theorem 2, the proposed data-driven estimator for the optimal scale parameter is consistent even when  $\sigma$  is unknown. Next, we show that the D PTR method, as specified in Algorithm 2, with  $\hat{\tau}_0$  and  $\hat{\beta}(\mathcal{S}, \mathcal{M}) = \hat{\beta}^*$ , will generate the same expected reward per experiment as the baseline case where  $\sigma$ ,  $\sigma_0$ , and  $\tau_0$  were known. We now introduce the expected per-experiment reward of the D PTR method with the estimated variance  $\hat{\sigma}^2 = \frac{1}{K} \sum_k s_k^2$ :

$$\bar{r}(\beta, \tau) = \frac{1}{K} \sum_{\substack{\hat{\tau}_k > \frac{N}{N+\beta} \frac{2\hat{\sigma} z_{1-\alpha/2}}{\sqrt{N}}} \tau_k.$$

**THEOREM 3.** Suppose Assumption 1 holds. As  $K \uparrow \infty$ , we have  $\bar{r}(\hat{\beta}^*, \hat{\tau}_0) \rightarrow_p \mathcal{R}(\beta^*, \tau_0)$ .

Combining Theorem 1 and Theorem 3 implies that the D PTR method with data-driven parameters can achieve an even higher reward than the IHT method with known  $\sigma$ , as long as the number of experiments  $K$  is sufficiently large. The key driving force behind this result is the delicate balance between bias and variance achieved by the D PTR method. The anchor  $\hat{\tau}_0$  leverages pooled data from a large number of experiments, significantly reducing variance compared to the individual signal  $\hat{\tau}_k$ . At the same time, our proposed D PTR method carefully controls bias through the optimally chosen scale parameter  $\hat{\beta}^*$ , ensuring an effective trade-off between bias and variance. Our method not just improves the estimation efficiency but also specifically optimizes the roll-out decision to maximize the expected reward.

Another commonly adopted approach to pool data from multiple experiments is the Bayesian method (Abadie et al. 2023). In Scenario 1, the ATE of experiment  $k$ ,  $\tau_k$ , follows the prior distribution  $\mathcal{N}(\tau_0, \sigma_0^2)$ . The outcome  $Y_k$  is sampled from

$$Y_{k,i} \sim \mathcal{N}(a_k + \tau_k D_{k,i}, \sigma_k^2), \quad i = 1, \dots, N.$$

For each experiment  $k$ , the platform randomly assigns  $N/2$  to the treatment condition and  $N/2$  to the control condition. Direct application of the Bayes rule implies the posterior distribution of  $\tau_k$  given the DM estimator  $\hat{\tau}_k$ :

$$\tau_k | \hat{\tau}_k \sim \mathcal{N}\left(\hat{\tau}_k \frac{N}{N + \beta_k^{\text{bayes}}} + \tau_0 \frac{\beta_k^{\text{bayes}}}{N + \beta_k^{\text{bayes}}}, \frac{N}{N + \beta_k^{\text{bayes}}} \frac{4\sigma_k^2}{N}\right), \quad \text{where } \beta_k^{\text{bayes}} = \frac{4\sigma_k^2}{\sigma_0^2}.$$

Similarly, in the data-driven setting, we apply the data pooling technique to estimate the prior mean  $\tau_0$  and variance  $\sigma_0^2$ . As shown in the proof of Theorem 2, we can obtain the same unbiased estimators for  $\tau_0$  and  $\sigma_0^2$ , i.e.,  $\hat{\tau}_0 = \frac{1}{K} \sum_{k=1}^K \hat{\tau}_k$  and  $\hat{\sigma}_0^2 = \frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k 4s_k^2$  where the  $s_k^2$  is the unbiased estimator for  $\sigma_k^2$  defined in Theorem 2. Then the data-driven scale parameter  $\hat{\beta}_k^{\text{bayes}}$  can be written as:

$$\hat{\beta}_k^{\text{bayes}} = \frac{4s_k^2}{\frac{1}{K} \sum_{k'} (\hat{\tau}_{k'} - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_{k'} 4s_{k'}^2}. \quad (21)$$

Comparing the scale parameters in Eqn. (20) and Eqn. (21) reveals insights on how our D PTR method differs from the Bayesian method. First, in the Bayesian method, the shrinkage parameter varies across different experiments, while in our method, it remains uniform, enhancing the effect of pooling data from different experiments. Second, the shrinkage parameter in our method includes an additional term that accommodates the significance level  $\alpha$ , making it decision-aware. To empirically compare the D PTR method with the Bayesian method, we follow Table 6 in Raftery (1995) and assumes the platform rolls out treatment  $k$  if the posterior probability of  $\tau_k > 0$  is at least  $1 - \alpha/2$ , in line with a two-sided test with significance level  $\alpha$  in the frequentist framework. The numerical experiments and empirical applications are presented in detail in Sections 5 and 6, respectively.

#### 4.2. Model with Covariates

In this subsection, we prove the D PTR method yields a higher expected reward than IHT for the OLS model with covariates. In this setting, based on Eqn. (6), the ATE estimator of treatment  $k$  is given by

$$\hat{\tau}_k = \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \mathbf{Y}_k = \tau_k + \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \boldsymbol{\epsilon}_k.$$

Under Assumption 1,  $\hat{\tau}_k$  follows a normal distribution with mean  $\tau_k$  and variance  $\sigma^2 b_k^2 / N$ , where  $b_k = \sqrt{N \mathbf{I}^\top (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{I}}$ . By Algorithm 2, we construct the new ATE estimator  $\bar{\tau}_k = \frac{N}{N + \beta(b_k)} \hat{\tau}_k + \frac{\beta(b_k)}{N + \beta(b_k)} \tau$  parametrized by scale  $\beta(b_k)$  and anchor  $\tau$ . The D PTR method then determines whether policy  $k$  will be rolled out based on  $\bar{\tau}_k$ . Different from the setting without covariates, the scale parameter depends on the treatment and covariate vector  $\mathbf{t}_k$  through the parameter  $b_k$ .

Similar to the model without covariates, we first assume that the variance  $\sigma^2$  is known. The platform uses  $z$ -statistic to construct the confidence bounds with the method  $\mathcal{M}(\cdot)$ . We define the per-experiment reward in this setting:

$$\tilde{r}(\beta(\cdot), \tau) = \frac{1}{K} \sum_{\bar{\tau}_k > \frac{N}{N + \beta(b_k)} \frac{\sigma b_k}{\sqrt{N}} z_{1-\alpha/2}} \tau_k. \quad (22)$$

It follows that  $\tilde{r}(0, \cdot) = \tilde{r}(0, \tau_0) = \hat{r}_{\text{IHT}}$ . Define  $\tilde{\mathcal{R}}(\beta(\cdot), \tau) := \lim_{K \uparrow \infty} \mathbb{E}[\tilde{r}(\beta(\cdot), \tau)]$  as the expected per-experiment reward when  $\tau = \tau_0$ . Our analysis begins with identifying the optimal scale parameter  $\beta(\cdot)$  that maximizes the expected per-experiment reward when  $\tau = \tau_0$  and  $K \rightarrow \infty$ .

**THEOREM 4.** Suppose Assumption 1 holds and  $\sigma$  is known. If  $\tau = \tau_0$ , and the optimal scale function  $\beta(\cdot)$  is:

$$\beta^*(b_k) := \arg \max_{\beta \geq 0} \tilde{\mathcal{R}}(\beta, \tau_0) = \frac{\sigma^2 b_k^2}{\sigma_0^2} + \frac{\sqrt{N} z_{1-\alpha} \sigma b_k}{\tau_0}.$$

In particular,  $\tilde{\mathcal{R}}(\beta^*(\cdot), \tau_0) \geq \tilde{\mathcal{R}}(0, \tau_0) = \lim_{K \uparrow +\infty} \mathbb{E}[\hat{r}_{IHT}]$ .

Theorem 4 characterizes the optimal scale parameter function  $\beta^*(\cdot)$  that maximizes the expected per-experiment reward of the DPTR roll-out method when the anchor is set at  $\tau = \tau_0$ . Unlike in Theorem 1, the scale parameter in this setting varies across experiments and depends on  $b_k$ . This dependency ensures that the scale parameter effectively leverages the diverse heterogeneous information ( $b_k^2$ ) obtained from different experiments. The sensitivity analysis for the parameters  $\sigma^2$ ,  $\sigma_0^2$ , and  $\tau_0$  remains the same as in the discussion following Theorem 1. Readers may refer to the previous subsection for details.

Similarly, in practice, the parameters  $\tau_0$ ,  $\sigma^2$ , and  $\sigma_0^2$  are unobservable by the platform. Hence, we estimate these parameters with the pooled data from all experiments and derive the estimator for function  $\beta^*(\cdot)$ . With estimated variance  $\hat{\sigma}^2 = \frac{1}{K} \sum_k s_k^2$ , we define the expected per-experiment reward under the DPTR method:

$$\bar{r}(\beta(\cdot), \tau) = \frac{1}{K} \sum_{\substack{\bar{r}_k > \frac{N}{N+\beta(b_k)} \frac{\hat{\sigma} b_k}{\sqrt{N}} z_{1-\alpha/2}}} \tau_k.$$

**THEOREM 5.** Suppose Assumption 1 holds. Define  $s_k^2 := \frac{1}{N-2-d_x} (\mathbf{Y}_k - \mathbf{t}_k (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \mathbf{Y}_k)^\top (\mathbf{Y}_k - \mathbf{t}_k (\mathbf{t}_k^\top \mathbf{t}_k)^{-1} \mathbf{t}_k^\top \mathbf{Y}_k)$ ,  $\hat{\tau}_0 = \frac{1}{K} \sum_k \hat{\tau}_k$  and,

$$\hat{\beta}^*(b_k) := \frac{b_k^2 (\frac{1}{K} \sum_{k'} s_{k'}^2)}{\frac{1}{K} \sum_{k'} (\hat{\tau}_{k'} - \hat{\tau}_0)^2 - \frac{1}{N} \left( \frac{1}{K} \sum_{k'} s_{k'}^2 \cdot \frac{1}{K} \sum_{k'} b_{k'}^2 \right)} + \frac{z_{1-\alpha/2} b_k \sqrt{N \frac{1}{K} \sum_{k'} s_{k'}^2}}{\hat{\tau}_0}.$$

We have, as  $K \uparrow \infty$ ,  $\hat{\tau}_0 \rightarrow_p \tau_0$  and for any  $b_k$ ,  $\hat{\beta}^*(b_k) \rightarrow_p \beta^*(b_k)$ . Furthermore, we have,  $\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) \rightarrow_p \mathcal{R}(\beta^*(\cdot), \tau_0)$ .

Theorem 4 and Theorem 5 together prove that, when the number of experiments  $K$  is sufficiently large, the DPTR method outperform the IHT method with covariates under the OLS specification.

## 5. Synthetic Experiments

In this section, we demonstrate the advantage of our proposed DPTR method over commonly adopted benchmarks such as the traditional IHT method and the Bayesian method using synthetic experiments across various scenarios outlined in Section 3.1 to Section 3.4. All relevant code can be found at GitHub.<sup>3</sup>

<sup>3</sup> See <https://github.com/shoucheng666/Data-Pooling-Treatment-Roll-Outs>.

We introduce two key metrics to evaluate the our methods: Optimality Ratio (OR) and Value of Data Pooling (VDP). OR measures the relative performance of a roll-out method (e.g, DPTR, Bayesian method or IHT) compared to the oracle per-experiment reward  $r^*$ , while VDP quantifies the relative reward improvement of DPTR or Bayesian method over IHT. Formally, we define:

$$\text{Optimality Ratio (OR) of Method } z = \frac{\hat{r}_z}{r^*}, \text{ Value of Data Pooling (VDP)} = \frac{\hat{r}_z}{\hat{r}_{\text{IHT}}} - 1,$$

where  $\hat{r}_z$  is the per-experiment reward generated by method  $z$ .

To provide a more comprehensive evaluation of our proposed method, we also frame the treatment roll-out decision problem as a classification task. Specifically, for each experiment  $k$ , if  $\tau_k > 0$ , it is labeled as a positive case; otherwise, it is labeled as a negative case. This classification perspective allows us to analyze DPTR and IHT as different classification algorithms. Thus, we further evaluate both methods using four standard classification metrics: Accuracy, Recall, Specificity, and Precision, each derived from the confusion matrix. These metrics are formally defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad \text{Specificity} = \frac{TN}{TN + FP}, \quad \text{Precision} = \frac{TP}{TP + FP},$$

where TP (True Positives) denotes the number of correctly identified positive cases; TN (True Negatives) denotes the number of correctly identified negative cases; FP (False Positives) denotes the number of negative cases incorrectly classified as positive; FN (False Negatives) denotes the number of positive cases incorrectly classified as negative. Unless otherwise specified, all experiments in this paper use a default significance level of  $\alpha = 0.05$ .

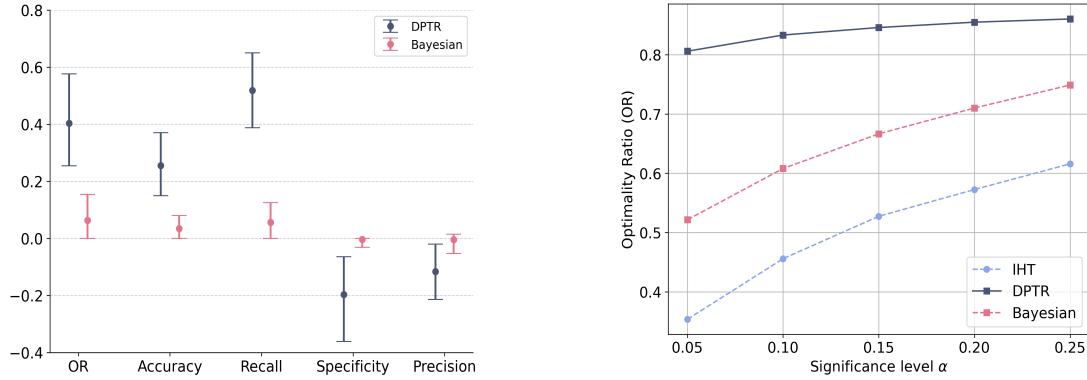
## 5.1. Non-overlapping Experiments and Linear Specification

When experiments are non-overlapping and model specifications are linear, results in Section 4 have already theoretically demonstrated how the DPTR method outperforms the IHT method by effectively balancing the bias-variance tradeoff in experiment roll-out decisions. In this subsection, we use synthetic experiments to illustrate the substantial edge of our proposed method even when the number of experiments is small or moderate.

**5.1.1. Without Covariate Information.** We set up the basic experimental setting as follows: the platform conducts  $K = 100$  experiments. The ground-truth ATE,  $\tau_k$ , is randomly sampled from a normal distribution  $\mathcal{N}(1, 3^2)$ . Each experiment has  $N = 10$  observations. The noise term  $\epsilon_{k,i}$  follows a normal distribution,  $\epsilon_{k,i} \sim \mathcal{N}(0, 3^2)$ .<sup>4</sup> Hereafter, we refer to this configuration of ATE

<sup>4</sup>In reality, each experiment typically contains a much larger number of observations, at the magnitude of hundreds of thousands or even millions, for a large-scale online platform (Kohavi et al. 2020). In this case, due to significant heterogeneity among users,  $\sigma$  is also orders of magnitude higher than  $\sigma_0$ . In our experiments, we proportionally scale down  $N$  and  $\sigma$ , while still capturing the key characteristics of the real-world scenario with reduced computational burden.

and outcome both following normal distributions as the normal-normal setting. We first compare DPTR (Bayesian method) and IHT with respect to different metrics: OR, Accuracy, Recall, Specificity, and Precision. We simulate 1,000 iterations and apply both methods in each iteration for the roll-out decision.



(a) Metric differences with 95% middle results: (b) Performance comparisons for different significance level  $\alpha$  DPTR (Bayesian method) minus IHT.

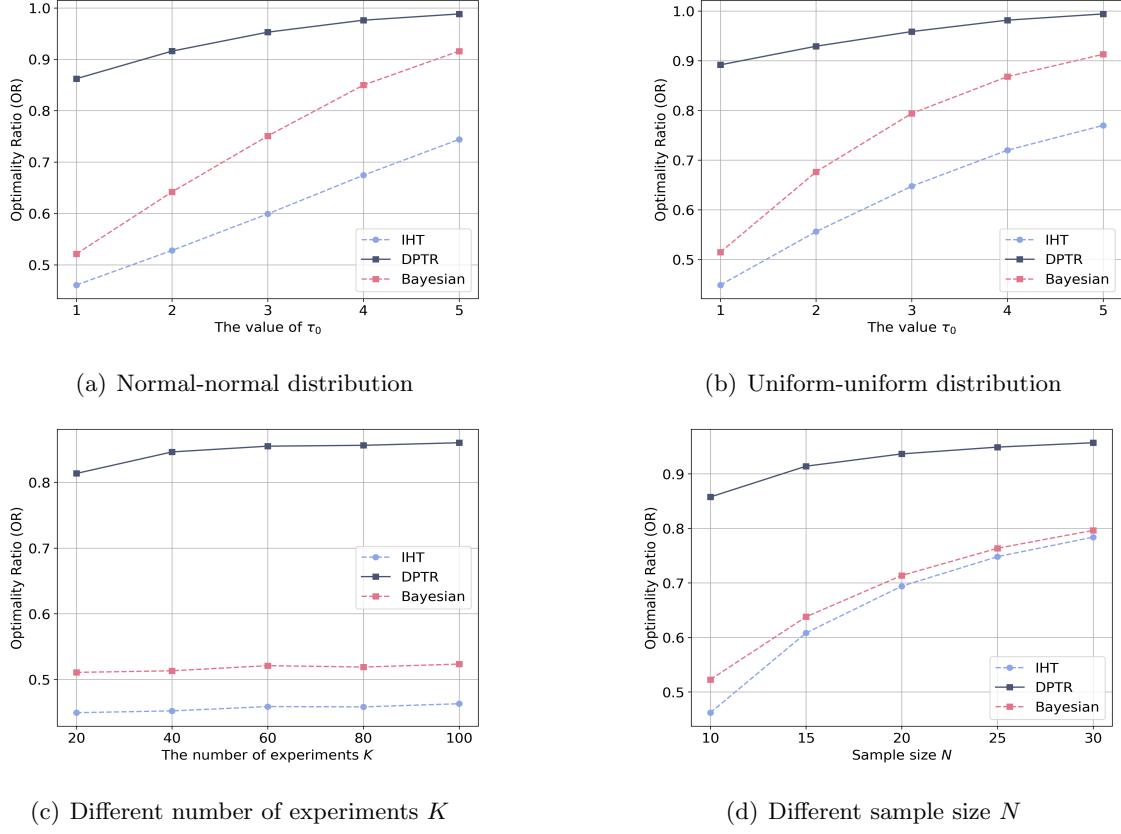
**Figure 2 Performance comparisons for non-overlapping experiments without covariates.**

Figure 2(a) depicts the intervals that cover the middle 95% of the differences in the five metrics between the DPTR (Bayesian method) and IHT methods across 1,000 iterations. A positive difference indicates that DPTR (Bayesian method) outperforms IHT, and vice versa. We find that DPTR and Bayesian methods consistently generate higher rewards than IHT, suggesting the superior performance of data pooling methods in roll-out decisions. A more careful look at the performance metrics reveals that the DPTR and Bayesian methods could make correct experiment roll-out decisions with a higher chance. Furthermore, DPTR and Bayesian methods significantly improve the recall. Compared to IHT, they can better identify the experiments that should be rolled out. On the other hand, such improvement is also at the cost of lower specificity and precision. This is because DPTR and Bayesian methods may mistakenly roll out some experiments with a negative treatment effect.

We compare the performance of the DPTR method with the Bayesian method. First, we observe that both data-pooling methods exhibit similar trends across all metrics, indicating that the core ideas and insights behind DPTR method are closely aligned with those of the Bayesian method. Second, the DPTR method achieves greater improvements in the optimality ratio, accuracy, and recall, albeit at the cost of slightly lower specificity and precision. This trade-off arises because our method is more decision-aware and places a stronger emphasis on maximizing reward, as discussed in Section 4.1.

It is useful to investigate the robustness of our method with respect to different significance levels  $\alpha$ . Specifically, we vary  $\alpha$  from 0.05 to 0.25 in increments of 0.05 and plot the OR metric for

DPTR, IHT and the Bayesian method. As shown in Figure 2(b), our DPTR method consistently achieves higher OR values across all tested  $\alpha$ 's and, notably, the performance gap remains significant regardless of the significance levels. Therefore, our proposed method strikes a delicate balance between statistical power and decision quality.

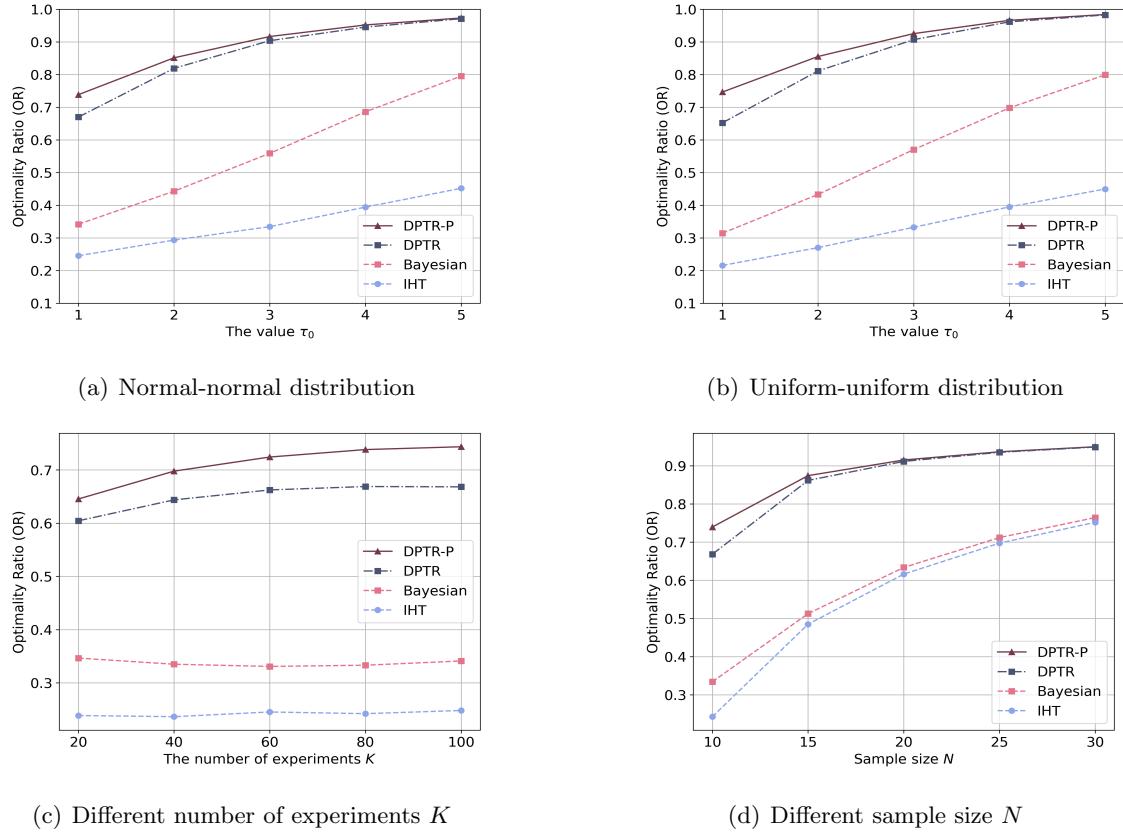


**Figure 3 Performance comparisons for non-overlapping experiments without covariates**

We proceed to further examine the robustness of DPTR under varying distributions, numbers of experiments and sample sizes. As shown in Figure 3, DPTR outperforms the IHT and Bayesian benchmarks regardless of prior ATE mean  $\tau_0$ , the distributions of prior ATE and noise terms, the number of experiments  $K$ , and the sample size  $N$ . This sensitivity analysis reveals our proposed method is particularly effective when the prior ATE mean  $\tau_0$  is small, the sample size  $N$  is small, or the number of experiments  $K$  is large. In these cases, DPTR assigns a high weight on the anchor  $\hat{\tau}_0$  estimated from data of multiple experiments, fully leveraging the benefit of data pooling. In particular, the finding related to sample size  $N$  provides a theoretical explanation for the empirical evidence presented in Chen et al. (2024), which shows that incorporating aggregate market information benefits small retailers more than large ones, likely due to the limited data available to smaller retailers.

**5.1.2. With Covariate Information.** We set up the experimental setting as follows: the number of covariates is set to  $d_x = 4$ , with covariate values sampled from  $x_{k,i}^j \sim U(0, 1)$ . The intercept  $a_k$  and the coefficients  $\theta_k$  are randomly drawn from  $U(-0.3, 0.5)$ . The number of experiments ( $K$ ), the sample size ( $N$ ), and the distributions of  $\tau_k$  and noise  $\epsilon_{k,i}$  are the same as in the previous subsection on non-overlapping experiments without covariates.

In this subsection, we test two variations of our method against the IHT and Bayesian benchmarks: (a) DPTR method using the common scale parameter  $\hat{\beta}$  defined in Eqn. (8), and the DPTR-P method using personalized scale parameters  $\hat{\beta}_k$  defined in Eqn. (9). Similar to the case without covariate information, we focus on comparing OR of different methods. The results are presented in Figure 4.



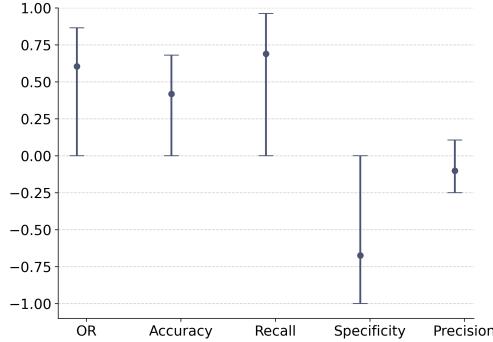
**Figure 4 Performance comparisons for non-overlapping experiments with covariates**

On one hand, the results are consistent with those in the case without covariates shown in Figure 3, demonstrating the robustness of our proposed method when incorporating covariate information. On the other hand, we further show that DPTR-P could achieve an even higher performance than DPTR with a shared scale parameter, which is well aligned with our theoretical results (Theorems 4 and 5), demonstrating the value of leveraging personalized information.

## 5.2. Non-overlapping Experiments and Non-linear Specification

In this subsection, we conduct a series of numerical experiments to evaluate the performance of DPTR under non-overlapping experiments and non-linear specification, as introduced in Section 3.2. We consider the following experimental setup: the platform runs  $K = 100$  experiments, each with  $N = 100$  observations. The noise  $\epsilon_{k,i}$  follows a normal distribution,  $\epsilon_{k,i} \sim \mathcal{N}(0, 3^2)$ . We set the number of covariates as  $d_x = 4$ , with covariate values sampled from  $x_{k,i}^j \sim U(0, 1)$ . The ground-truth response function is defined as  $g_k(X_{k,i}) = [\gamma_{k,0}^\top X_{k,i}, \gamma_{k,1}^\top X_{k,i}]^\top$ , where the components of  $\gamma_{k,0}$  and  $\gamma_{k,1}$  are independently drawn from the distribution  $U(-0.3, 0.5)$ . To estimate the nuisance parameter, we use a two-layer fully connected neural network with 10 units per layer and ReLU activations, without dropout, as the function class  $\mathcal{F}$ . As illustrated in Section 3.2, double machine learning and cross-fitting techniques are applied to estimate the ATE of each experiment  $k$ , denoted by  $\hat{\tau}_k$ .

Similar to Section 5.1.1, we also run the simulation for 1,000 iterations and show the middle 95% of the differences in five metrics between DPTR and IHT in the five metrics (OR, Accuracy, Recall, Specificity and Precision), as shown in Figure 5. It is clearly illustrated in Figure 5 that the results are consistent with those of DPTR method in the setting of non-overlapping experiments and non-linear specification (see Figure 2(a)).



**Figure 5** DPTR vs. IHT: Non-overlapping experiments and non-linear specification

Next, we incorporate the personalized scale parameter  $\hat{\beta}_k$  defined in Eqn. (13), and evaluate the performance of our methods under two scenarios: one with a relatively large number of experiments ( $K = 100$ ), and the other with a relatively small number ( $K = 5$ ). Under each situation, we vary  $\sigma$  from 1 to 5 in increments of 1. For each parameter specification, we run the simulation for 1,000 iterations and report the average OR of each roll-out method in Table 1. Our simulation results show that DPTR and DPTR-P consistently yield significant reward improvements over IHT, regardless of the number of experiments  $K$  and noise variance  $\sigma^2$ . The improvement is noticeably greater when the number of experiments is larger. In this case, the estimation accuracy of  $\hat{\beta}$ ,  $\hat{\beta}_k$ , and  $\hat{\tau}_0$  is higher, rendering data pooling via our methods more effective. We also observe that DPTR-P consistently outperforms DPTR, except for the case where  $K$  is large and  $\sigma^2$  is small.

Even though  $b_k^2$  is constructed based on a misspecified linear model, it still provides benefits when the number of experiments is small or the variance is large.

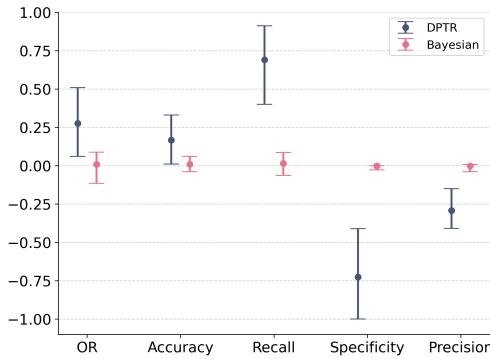
| Noise Variance ( $\sigma^2$ ) | $K = 100$ |        |        | $K = 5$ |        |        |
|-------------------------------|-----------|--------|--------|---------|--------|--------|
|                               | IHT       | DPTR   | DPTR-P | IHT     | DPTR   | DPTR-P |
| $1^2$                         | 0.4097    | 0.9214 | 0.8930 | 0.3804  | 0.6768 | 0.7476 |
| $2^2$                         | 0.1708    | 0.8394 | 0.8417 | 0.1662  | 0.4855 | 0.6728 |
| $3^2$                         | 0.1003    | 0.7123 | 0.7612 | 0.0990  | 0.3726 | 0.5882 |
| $4^2$                         | 0.0745    | 0.5831 | 0.7029 | 0.0739  | 0.3455 | 0.5357 |
| $5^2$                         | 0.0596    | 0.5349 | 0.6858 | 0.0468  | 0.2982 | 0.4951 |

**Table 1 Performance comparison under OR: Non-overlapping experiments and non-linear specification**

### 5.3. Overlapping Experiments and Linear Specification

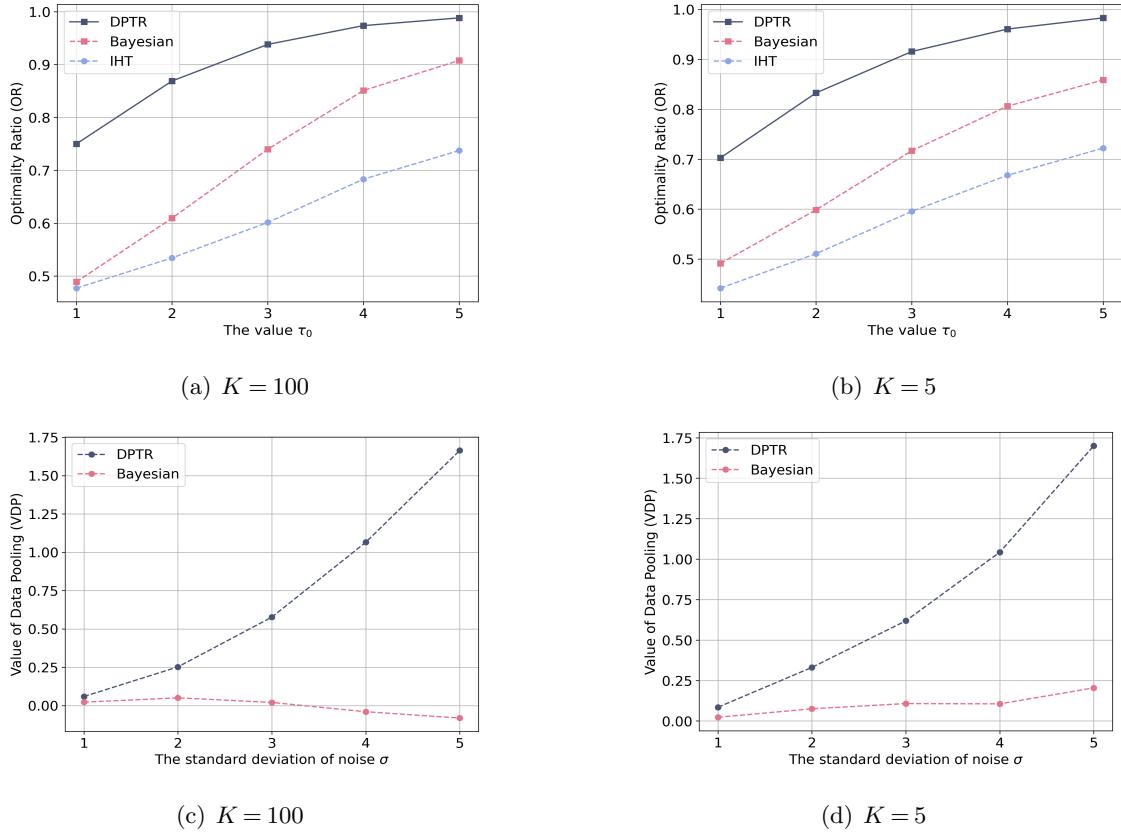
We now numerically test the performance of DPTR under overlapping experiments and linear specification, as described in Section 3.3. The experimental setup is as follows: the platform runs  $K = 100$  experiments, with  $N = 10 + K$  observations. The  $D_{k,i}$ 's are i.i.d. Bernoulli random variables with  $\mathbb{P}[D_{k,i} = 1] = \mathbb{P}[D_{k,i} = 0] = 0.5$ . The noise  $\epsilon_{k,i}$  follows a normal distribution,  $\epsilon_{k,i} \sim \mathcal{N}(0, 3^2)$ . The ground-truth ATE,  $\tau_k$ , is randomly sampled from a normal distribution  $\mathcal{N}(1, 3^2)$ .

We repeat the experiment 1,000 times using the Bayesian, DPTR and IHT methods, and present the middle 95% of the differences in five metrics (OR, Accuracy, Recall, Specificity, and Precision) to compare their performance, as shown in Figure 6. The results closely mirror those in Figure 2 (a), even with overlapping experiments.



**Figure 6 DPTR (Bayesian method) vs. IHT: Overlapping experiments and linear specification**

We compare DPTR, IHT, and Bayesian methods and report OR and VDP in Figure 7. The findings are closely aligned with those reported in Section 5.1. The results show that our DPTR method remains competitive even when the number of experiments is small. In summary, the proposed DPTR method has a robust performance under linear specification even when users are treated by multiple experiments simultaneously.



**Figure 7 Performance comparisons under overlapping experiments and linear specification**

#### 5.4. Overlapping Experiments and Non-linear Specification

In this subsection, we test the performance of our method in the scenario introduced in Section 3.4. We evaluate the performance of our method only when  $K$  is relatively small. The experimental setup is as follows: the platform runs  $K = 5$  experiments, with  $N = 100 + K$  observations. The noise  $\epsilon_{k,i}$  follows a normal distribution,  $\epsilon_{k,i} \sim \mathcal{N}(0, 3^2)$ . We set the number of covariates as  $d_x = 4$ , the covariate distribution as  $x_{k,i}^j \sim U(0, 1)$ , and the true response function as  $g(X_i) = [\gamma_0^\top X_i, \gamma_1^\top X_i, \dots, \gamma_K^\top X_i]^\top$ , where the coefficients  $\gamma_0, \gamma_1, \dots, \gamma_K$  are randomly drawn from the distribution  $U(-0.3, 0.5)$ . The estimation is also based on double machine learning and cross-fitting. To estimate the nuisance parameter  $g(\cdot)$ , we adopt the two-layer fully connected neural network with  $K + 10$  units per layer and ReLU activations, without dropout, to estimate  $g(\cdot)$ . As reported in Table 2, D PTR consistently outperforms I HT, demonstrating that our proposed data pooling technique effectively combines data from multiple experiments even under overlapping experiments and nonlinear specification.

#### 5.5. Robustness under Model Misspecification

Our analysis so far has focused on the assumption that the policy ATEs are linearly additive. In practice, however, this assumption does not hold in general (Ye et al. 2025). To understand how

| Noise Variance ( $\sigma^2$ ) | IHT (OR) | DPTR (OR) |
|-------------------------------|----------|-----------|
| 1 <sup>2</sup>                | 0.3590   | 0.6897    |
| 2 <sup>2</sup>                | 0.1499   | 0.5257    |
| 3 <sup>2</sup>                | 0.0992   | 0.5008    |
| 4 <sup>2</sup>                | 0.0650   | 0.4438    |
| 5 <sup>2</sup>                | 0.0512   | 0.4344    |

**Table 2** The comparison of OR under different  $\sigma^2$ .

well our proposed data pooling method works when the treatment effects of different policies are not linearly additive, we consider the *Generalized Sigmoid Form II* DGP in [Ye et al. \(2025\)](#):

$$Y_i = \frac{v}{1 + \exp(-g(X_i)^\top \mathbf{t}_i)} + \epsilon_i, \quad i = 1, \dots, N, \quad (23)$$

where  $g(\cdot) : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{K+1}$  is the true response function,  $\mathbf{t}_i$  is the treatment vector which includes a constant term, and  $\epsilon_i$  denotes the i.i.d. random noise. Thus, the optimal reward the platform can obtain is given by:

$$r^* = \max_{\mathbf{t}} \mathbb{E}[Y|\mathbf{t}] - \mathbb{E}[Y|\mathbf{t}_0],$$

where  $\mathbf{t}_0$  is the base treatment vector in which the treatment indicators for all experiments equal to zero. While the data generating process follows Eqn. (23), we deliberately ignore the non-linear model specifications and apply the same method in Section 5.4 to decide whether the experiment  $k$  should be implemented. After implementing Algorithm 1 and 2, we obtain the roll-out decisions  $\hat{\mathcal{A}}_{\text{IHT}}$  and  $\hat{\mathcal{A}}_{\text{DPTR}}$ . The reward obtained by the IHT and DPTR can be written as:

$$\hat{r}_{\text{IHT}} = \mathbb{E}[Y|\hat{\mathcal{A}}_{\text{IHT}}] - \mathbb{E}[Y|\mathbf{t}_0], \quad \bar{r}(\beta, \tau) = \mathbb{E}[Y|\hat{\mathcal{A}}_{\text{DPTR}}] - \mathbb{E}[Y|\mathbf{t}_0].$$

The experimental setup is as follows: the error term is sampled from a normal distribution  $\mathcal{N}(0, 3^2)$ , and the number of experiments is set to  $K = 4$ . We draw  $v$  from the uniform distribution  $U(10, 20)$ , and define the function  $g(X_i)$  as  $g(X_i) = \{\gamma_0^\top X_i, \gamma_1^\top X_i, \dots, \gamma_K^\top X_i\}$  where each vector  $\gamma_0, \gamma_1, \dots, \gamma_K$  consist of components which are independently drawn from  $U(-0.3, 0.5)$ . The covariate  $X_i$  is of dimension  $d_x = 4$ , where each component is sampled independently from  $U(0, 1)$ . To provide a more comprehensive comparison of the performance of the IHT and DPTR methods, we vary the number of experiments from 4 to 7 with an increment of 1, and also vary  $\sigma$  from 3 to 5 in increments of 1. We repeat each setting 1000 times and report the average OR values in Table 3.

| K | $\mathcal{N}(0, 3^2)$ |        | $\mathcal{N}(0, 4^2)$ |        | $\mathcal{N}(0, 5^2)$ |        |
|---|-----------------------|--------|-----------------------|--------|-----------------------|--------|
|   | IHT                   | DPTR   | IHT                   | DPTR   | IHT                   | DPTR   |
| 4 | 0.0269                | 0.3754 | 0.0259                | 0.3862 | 0.0217                | 0.3862 |
| 5 | 0.0142                | 0.4267 | 0.0151                | 0.3913 | 0.0104                | 0.4043 |
| 6 | 0.0065                | 0.4641 | 0.0057                | 0.4346 | 0.0058                | 0.4301 |
| 7 | 0.0023                | 0.5133 | 0.0013                | 0.4864 | 0.0032                | 0.4945 |

**Table 3** The performance comparison with OR under different K and variance of error term.

First, we observe that the DPTR method consistently outperforms the IHT method, regardless of the number of experiments or the magnitude of the error term. Second, due to the effect of nonlinearity, the performance of both the DPTR and IHT methods does not show a strictly monotonic decline as the variance of error term increases. Finally, an interesting phenomenon is that when nonlinearity is ignored, increasing the number of experiments tends to amplify the degree of nonlinearity. This, in turn, leads to deteriorating performance of the IHT method, while the DPTR method continues to improve. This finding further demonstrates that the DPTR method is capable of rolling out high-reward treatments even under model misspecification, highlighting its robustness.

## 6. Applications to Real-world A/B Tests

In this section, we evaluate the performance of the DPTR method using the real-world A/B testing data, covering both non-overlapping and overlapping scenarios. Section 6.1 reports experiment results based on a publicly available dataset from Criteo (Diemert et al. 2018) to demonstrate the performance of our proposed method in the non-overlapping scenario. In Section 6.2, we utilize the experimental data from Ye et al. (2025), which includes multiple experiments, to demonstrate the DPTR method’s effectiveness in the overlapping scenario.

### 6.1. Non-Overlapping A/B Testing

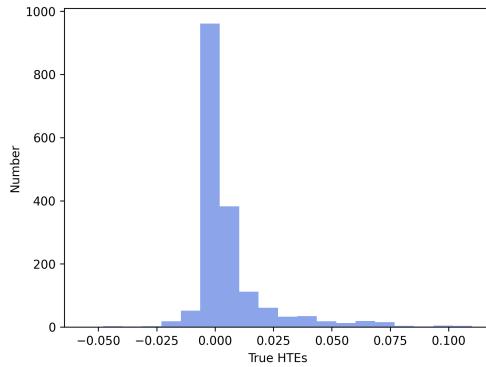
In this subsection, we will evaluate the performance of the DPTR method in a non-overlapping scenario. The dataset<sup>5</sup> used in this analysis originates from a randomized controlled trial (RCT), conducted by Criteo, an advertising platform, as part of a large-scale randomized ad-targeting campaign. In this RCT, a randomly selected portion of the population was deliberately excluded from being targeted by advertisements. This RCT was initially released to benchmark uplift modeling methods (Diemert et al. 2018) with “visits” as the outcome of interest. The dataset comprises 13,979,592 rows, each representing a user characterized by 12 covariates, a treatment indicator for advertisement exposure, and a binary label indicating whether the user visited the advertised site. The treatment rate is 85% and the average visit rate is 4.70%.

Unlike the synthetic data setting explored in Section 5, this dataset is from a single experiment and contains user covariate information. Consequently, our focus shifts to determining customized treatment rollouts, specifically, deciding whether personalized recommendations should be offered to users. To validate the proposed approach using this dataset, we first group all samples based on user covariates. Then, we estimate the “true” HTEs across different groups using the full dataset. Next, we evaluate both DPTR and IHT methods, using random samples from the dataset associated with each group. The detailed validation procedure is as follows:

<sup>5</sup> The dataset can be accessed via the link <https://ailab.criteo.com/criteo-uplift-prediction-dataset/>.

1. Group Generation: We partition the dataset based on the medians of user covariates. Since the covariate information in this dataset is encrypted, we cannot group the data based on specific covariate values. Instead, we categorize each covariate into two groups based on its median value. As many covariates have a large number of values equal to the median, we randomly assign samples to ensure that both groups have equal sample sizes. Thus, we partition the entire dataset based on the combinations formed by the realizations of the 12 covariates. Since the features are not completely independent, the number of data points in the groups formed by splitting on the median of each feature is not necessarily the same. Thus, to ensure each group has enough data for reliable analysis, we exclude groups with fewer than 1,000 data entries. This results in 1,744 groups, that is,  $K = 1,744$ .

2. “True” HTEs Calculation: For group  $k$ , let  $D_{k,i} \in \{0,1\}$  represent the treatment variable for subject  $i$ , where  $D_{k,i} = 1$  indicates the implementation of the personalized recommendation, and  $D_{k,i} = 0$  indicates no implementation. Let  $Y_{k,i}$  denote the outcome, indicating whether the user visits the recommended advertisement. The “true” HTE for group  $k$  is then given by:  $\tau_k = \frac{1}{N_{k,1}} \sum_{D_{k,i}=1} Y_{k,i} - \frac{1}{N_{k,0}} \sum_{D_{k,i}=0} Y_{k,i}$  where  $N_{k,1}$  and  $N_{k,0}$  are the total number of subjects who experienced and did not experience the treatment, respectively. The histogram of the “true” HTEs across all groups, estimated from the entire dataset, is shown in Figure 8.

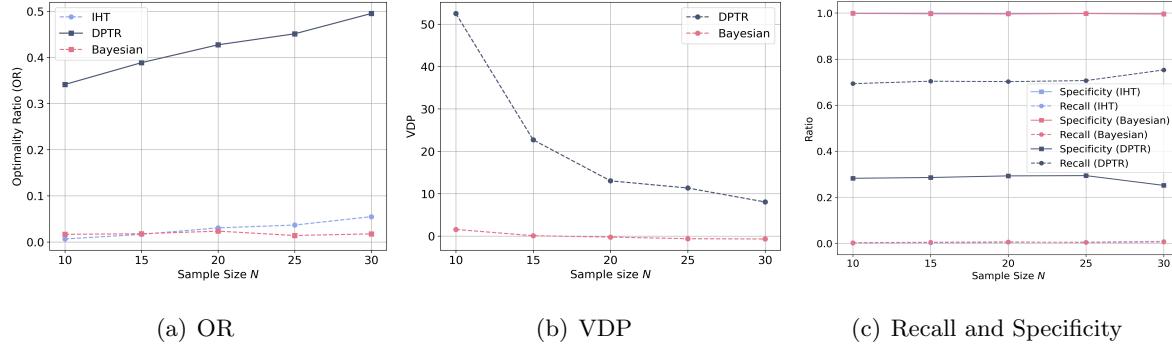


**Figure 8 Histogram of true HTEs for all groups for Criteo dataset.**

3. Random Sampling and Evaluation: For each group, we randomly select  $N$  users, with  $N/2$  users drawn from the treatment group ( $D_{k,i} = 1$ ) and  $N/2$  users drawn from the control group ( $D_{k,i} = 0$ ) for every  $k$ . Using this randomly sampled sub-dataset, we apply both the IHT and DPTR methods (Algorithms 1 and 2, respectively) to generate roll-out decisions. Here, we select the difference-in-mean method as  $\mathcal{M}(\cdot)$  and set the significance level  $\alpha = 0.05$ . Finally, we can compare their performance by measuring different metrics.

Since the sample sizes vary across groups, we normalize  $\tau_k$  when calculating the OR and VDP values by multiplying it with the normalizing factor  $\frac{N_{k,0}+N_{k,1}}{\sum_k(N_{k,0}+N_{k,1})}$ , which represents the proportion

of group  $k$ 's data size relative to the total data size of all groups. We vary the sample size  $N$  from 10 to 30 in increments of 5 and repeat the experiment 1,000 times. The averaged results are shown in Figure 9.



**Figure 9 Performance comparisons in different sample size  $N$  in Criteo dataset.**

According to Figure 9(a), the DPTR method consistently outperforms the IHT method, irrespective of the sample size  $N$ . Additionally, as the sample size increases, the performance difference between DPTR and IHT becomes larger. However, the Bayesian method performs almost on par with the IHT method. The significant performance difference between DPTR method and the Bayesian method hinges on how the shrinkage parameter  $\beta$  is constructed: while the one for DPTR requires the average sample variance across experiments (Eqn. 20), the one for the Bayesian method utilizes the individual-level sample variance (Eqn. 21). As reported in Figure 8, most true HTEs are centered around zero, which may benefit most from shrinkage, yet their sample variances tend to be relatively small. Thus, the associated shrinkage parameters  $\hat{\beta}_k^{\text{bayes}}$  are closer to zero, resulting in a similar performance of the IHT method. In contrast, DPTR leverages the average sample variance, which may significantly deviate from zero. This allows the shrinkage parameter to more effectively guide decision-making and achieve superior OR values. In sum, these results further highlight that our DPTR method is more decision-aware and robust than the Bayesian method.

Figure 9(b) further reveals that the relative performance advantage of the DPTR method over the IHT method increases significantly as the amount of experimental data decreases. This finding highlights that, when experimental data is limited, the DPTR method provides substantial performance improvements.

Figure 9(c) shows that when  $N$  is small, the Recall value under the IHT method and Bayesian method is closer to 0, indicating that few or no personalized recommendations are rolled out to specific groups, even when these groups may be associated with positive ATEs. Because few groups are treated with the personalized recommendation, the Specificity value is also higher under the IHT and Bayesian methods. On the other hand, despite large variation in the estimated ATEs across groups, the DPTR method mitigates this issue by pooling the estimates, which reduces variance

and shifts the estimate towards a more positive range. As a result, the Recall value under the DPTR method increases significantly. This comes at the cost of mistakenly selecting some groups for the personalized recommendation. However, by balancing Recall and Specificity, the DPTR method generates a much higher reward compared to the IHT and Bayesian methods when the personalized roll-out decisions are involved. Finally, we remark that additional evaluation of the DPTR method using Expedia’s experimental data in a similar non-overlapping setting is provided in Appendix B.

## 6.2. Overlapping A/B Tests

In this subsection, we will use the experimental data from Ye et al. (2025) to demonstrate the performance of our DPTR method in the overlapping scenario. The dataset was collected from a large-scale online short-video-sharing platform, which serves hundreds of millions of users globally each day.

The dataset comprises a unique set of three A/B tests or treatments, each of which examines the effect of a major adjustment to the video recommendation algorithm on one of three main pages of the online platform: (i) the Discover Page (DP), (ii) the Live Page (LP), and (iii) the For You Page (FYP). As with most A/B tests conducted on online platforms, the primary objective is to enhance user engagement, which is well approximated by the amount of screen time a user spends on the platform each day. Each experiment is randomized using a distinct hash function of user IDs, ensuring that the treatment assignment mechanisms across experiments are mutually independent. Ye et al. (2025) have demonstrated the presence of interaction effects across experiments in this dataset. As a result, the three experiments yield eight distinct treatment combinations. Using stratified sampling, Ye et al. (2025) construct a new dataset with approximately 258,325 users in each treatment combination. We refer to the total data as population data. A detailed description of the dataset and experiments can be found in Ye et al. (2025).

First, using the full population data, we compute the ground-truth ATEs for all treatment combinations. The relative ATE values for the eight treatment combinations are reported in Table 2 of Ye et al. (2025). For combinations with statistically insignificant ATEs, we set the ground-truth ATE to zero. Next, we randomly sample  $N$  users for each treatment combination from the corresponding population data to construct a new dataset, which we refer to as the historical experimental data. We then apply both the DPTR and IHT methods to this dataset to make roll-out decisions, and evaluate their performance using OR values calculated based on the ground-truth ATEs. This sampling procedure is repeated 1,000 times, and we report the average OR values across these iterations.

To more robustly demonstrate the advantage of the DPTR method over the IHT method, we vary  $N$  across the values 1000, 5000, 10000, and 20000. For the method  $\mathcal{M}$ , we consider the following

four methods: (i) DM, (ii) OLS without covariate information, (iii) OLS with covariate information, and (iv) DML. Next, we provide a detailed description of the implementation procedures for each method as follows:

(i) **DM**: For this method, we focus on a non-overlapping scenario without covariate information. Accordingly, we use only the data corresponding to the three treatment combinations:  $(1, 0, 0)$ ,  $(0, 1, 0)$ , and  $(0, 0, 1)$ . The model assumed for this method follows the form of Eqn. (2).

(ii) **OLS without covariate information**: For this method, we consider an overlapping scenario that utilizes data from all treatment combinations. However, covariate information is not included, and the assumed model follows the form of Eqn. (14).

(iii) **OLS with covariate information**: For this method, we consider an overlapping scenario and incorporate covariate information. The assumed model follows the form of Eqn. (16).

(iv) **DML** : For this method, we consider an overlapping scenario and use the DML approach to estimate the ATEs. We assume that the policy ATEs are linearly additive, and the model follows the form of Eqn. (17).

The averaged OR values of 1,000 instances for different methods  $\mathcal{M}$  and sample size  $N$  are reported in the Table 4. First, we observe that our DPTR method consistently outperforms the IHT method with different methods and sample sizes. This confirms the effectiveness of our method for prescribing roll-out decisions using real-world datasets. Second, for any given method  $\mathcal{M}$ , increasing the sample size  $N$  improves the performance for both the DPTR and IHT methods. This is consistent with our intuition: a larger sample size typically results in a smaller variance. Furthermore, for a fixed sample size, when adopting IHT method, we may observe that “OLS without covariate” outperforms “OLS with covariate”, which, in turn, results in higher ORs than DML. This may imply the potential risk of model misspecification. More importantly, regardless of whether model misspecification exists, DPTR method outperforms the IHT method, demonstrating its robustness, especially when model misspecification may arise in analyzing real-world datasets.

| Method ( $\mathcal{M}$ ) | Roll-out method | 1,000  | 5,000  | 10,000 | 20,000 |
|--------------------------|-----------------|--------|--------|--------|--------|
| DM                       | IHT             | 0.0207 | 0.0408 | 0.0584 | 0.0732 |
|                          | DPTR            | 0.1610 | 0.2482 | 0.2780 | 0.3080 |
| OLS without covariate    | IHT             | 0.0568 | 0.1070 | 0.2281 | 0.3564 |
|                          | DPTR            | 0.2878 | 0.4213 | 0.5257 | 0.6926 |
| OLS with covariate       | IHT             | 0.0161 | 0.0853 | 0.1499 | 0.2667 |
|                          | DPTR            | 0.4342 | 0.5213 | 0.5437 | 0.6060 |
| DML                      | IHT             | 0.0080 | 0.0398 | 0.1058 | 0.2145 |
|                          | DPTR            | 0.3654 | 0.4208 | 0.4845 | 0.5474 |

**Table 4** The comparison of OR under different methods  $\mathcal{M}$  and sample size  $N$ .

## 7. Conclusion

In conclusion, we introduce the Data-Pooling Treatment Roll-Out (DPTR) framework to improve decision-making in online experiments by aggregating data across multiple experiments. By addressing challenges like limited traffic, heterogeneous treatment effects, and overlapping experiments, DPTR effectively balances bias and variance in treatment effect estimation. Theoretical analysis, synthetic experiments, and empirical validation with Expedia data show that DPTR consistently outperforms traditional individual hypothesis testing (IHT), especially with small sample sizes and many experiments. This study highlights the value of data pooling for policy roll-outs, with potential future extensions to non-linear treatment effects and broader decision-making contexts.

In this paper, we focus on the case of binary treatments and propose a method that pools data across multiple experiments to improve the per-experiment reward by shrinking the estimators derived from individual datasets. Although our primary analysis is centered on binary treatments, the underlying idea is broadly applicable. For example, it can be extended to settings with continuous treatments, such as pricing interventions in A/B testing, where the goal is to estimate dose-response relationships (Zhang et al. 2025). Moreover, the approach can be applied to observational (non-experimental) data, where confounding may arise, offering a principled way to stabilize estimates in the presence of such complexities (Jiang and Li 2025, Chitla et al. 2025). Furthermore, our framework can be extended to optimization problems with uncertain objectives and prescribe roll-out decisions, potentially subject to constraints (Natarajan et al. 2011).

## References

- Abadie A, Agarwal A, Imbens G, Jia S, McQueen J, Stepaniants S (2023) Estimating the value of evidence-based decision making. [arXiv preprint arXiv:2306.13681](https://arxiv.org/abs/2306.13681).
- Amemiya T (1985) *Advanced econometrics* (Harvard university press).
- Angrist JD, Pischke JS (2009) *Mostly harmless econometrics: An empiricist's companion* (Princeton university press).
- Athey S, Chetty R, Imbens G (2025) Using experiments to correct for selection in observational studies. URL <https://arxiv.org/abs/2006.09676>.
- Athey S, Imbens G (2016) Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* 113(27):7353–7360, URL <http://dx.doi.org/10.1073/pnas.1510489113>.
- Bastani H (2021) Predicting with proxies: Transfer learning in high dimension. *Management Science* 67(5):2964–2984.
- Bojinov I, Simchi-Levi D, Zhao J (2023) Design and analysis of switchback experiments. *Management Science* 69(7):3759–3777.
- Booking (2019) The role of experimentation at booking.com. <https://partner.booking.com/en-us/click-magazine/industry-perspectives/role-experimentation-bookingcom>, retrieved on June 7, 2025.
- Box GE, Hunter WH, Hunter S, et al. (1978) *Statistics for experimenters*, volume 664 (John Wiley and sons New York).

- Brown LD (1971) Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *The Annals of Mathematical Statistics* 42(3):855–903.
- Candogan O, Chen C, Niazadeh R (2021) Correlated cluster-based randomized experiments: Robust variance minimization. *Chicago Booth Research Paper* (21-17).
- Caruana R (1997) Multitask learning. *Machine learning* 28:41–75.
- Chan T (2021) Embrace overlapping a/b tests and avoid the dangers of isolating experiments. <https://blog.statsig.com/embracing-overlapping-a-b-tests-and-the-danger-of-isolating-experiments-cb0a69e09d3>, retrieved on August 1, 2025.
- Chen Y, Cui X, Li A, Wu B, Yang L (2024) The role of digital platforms in data markets: How data sharing through advanced analytics empowers small business innovation. Available at SSRN 4878048 .
- Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* C1–C68.
- Cheung WC, Simchi-Levi D, Wang H (2017) Dynamic pricing and demand learning with limited price experimentation. *Operations Research* 65(6):1722–1731.
- Chitla S, Jagabathula S, Venkataraman A (2025) Improving demand prediction by accounting for unobserved factors. Available at SSRN 4244086 .
- Cui R, Li J, Zhang DJ (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science* 66(3):1071–1094.
- CXL (2020) Can you run multiple a/b tests at the same time? [https://cxl.com/blog/can-you-run-multiple-a-b-tests-at-the-same-time/?gad\\_source=1&gad\\_campaignid=22595223963&gbraid=AAAAAADD4CBfgn732DU6wl5src48M5r9kc&gclid=CjwKCAjwy7HEBhBJEiwA5hQNopKI8gUQZBcywPpG2B3hvAHpgT7RxnCsNfHi3y72ieNNA7dPfkYX8xoCC\\_gQAvD\\_BwE](https://cxl.com/blog/can-you-run-multiple-a-b-tests-at-the-same-time/?gad_source=1&gad_campaignid=22595223963&gbraid=AAAAAADD4CBfgn732DU6wl5src48M5r9kc&gclid=CjwKCAjwy7HEBhBJEiwA5hQNopKI8gUQZBcywPpG2B3hvAHpgT7RxnCsNfHi3y72ieNNA7dPfkYX8xoCC_gQAvD_BwE), retrieved on August 1, 2025.
- Dasgupta T, Pillai NS, Rubin DB (2015) Causal inference from 2k factorial designs by using potential outcomes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77(4):727–753.
- Diemert E, Artem B, Renaudin C, Massih-Reza A (2018) A large scale benchmark for uplift modeling. *Proceedings of the AdKDD and TargetAd Workshop, KDD, London, United Kingdom, August, 20, 2018* (ACM).
- Efron B, Morris C (1977) Stein’s paradox in statistics. *Scientific American* 236(5):119–127.
- Farrell MH, Liang T, Misra S (2020) Deep learning for individual heterogeneity: An automatic inference framework. *arXiv preprint arXiv:2010.14694* .
- Feng Q, Li L, Shanthikumar JG (2023) Transfer learning, cross learning and co-learning across newsvendor systems with operational data analytics (ODA). *working paper* .
- Gallino S, Karacaoglu N, Moreno A (2023) Need for speed: The impact of in-process delays on customer behavior in online retail. *Operations Research* 71(3):876–894.
- Greene WH (2003) *Econometric analysis* (Pearson education india).
- Gui GZ (2024) Combining observational and experimental data to improve efficiency using imperfect instruments. *Marketing Science* 43(2):378–391.
- Gupta V, Kallus N (2022) Data pooling in stochastic optimization. *Management Science* 68(3):1595–1615.
- Hahn PR, Murray JS, Carvalho CM (2020) Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis* 15(3):965–1056.
- Imbens G, Kallus N, Mao X, Wang Y (2025) Long-term causal inference under persistent confounding via data combination. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 87(2):362–388.

- Imbens GW, Rubin DB (2015) *Causal inference in statistics, social, and biomedical sciences* (Cambridge university press).
- Jiang Z, Li J (2025) Instrumenting while experimenting: An empirical method for competitive pricing at scale. *Operations Research*.
- Johari R, Li H, Liskovich I, Weintraub GY (2022) Experimental design in two-sided platforms: An analysis of bias. *Management Science* 68(10):7069–7089.
- Kohavi R, Deng A, Frasca B, Walker T, Xu Y, Pohlmann N (2013) Online controlled experiments at large scale. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1168–1176.
- Kohavi R, Tang D, Xu Y (2020) *Trustworthy online controlled experiments: A practical guide to a/b testing* (Cambridge University Press).
- Kohavi R, Thomke S (2017) The surprising power of online experiments. *Harvard business review* 95(5):74–82.
- Lada A, Peysakhovich A, Aparicio D, Bailey M (2019) Observational data for heterogeneous treatment effects with application to recommender systems. 199–213, EC '19 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450367929, URL <http://dx.doi.org/10.1145/3328526.3329558>.
- Lei D, Qi Y, Liu S, Geng D, Zhang J, Hu H, Shen ZJM (2024) Pooling and boosting for demand prediction in retail: A transfer learning approach. *Manufacturing & Service Operations Management*.
- Lewis RA, Rao JM (2015) The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics* 130(4):1941–1973.
- Luca M, Bazerman MH (2021) *The power of experiments: Decision making in a data-driven world* (Mit Press).
- Microsoft (2023) A/b interactions: A call to relax. <https://www.microsoft.com/en-us/research/articles/a-b-interactions-a-call-to-relax/>, retrieved on August 1, 2025.
- Nabi S, Nassif H, Hong J, Mamani H, Imbens G (2022) Bayesian meta-prior learning using empirical bayes. *Management Science* 68(3):1737–1755.
- Nandy P, Venugopalan D, Lo C, Chatterjee S (2021) A/b testing for recommender systems in a two-sided marketplace. *Advances in Neural Information Processing Systems* 34:6466–6477.
- Natarajan K, Teo CP, Zheng Z (2011) Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations research* 59(3):713–728.
- Ni T (2025) Decision analytics of switchback experiments: A robust optimization approach. [Available at SSRN 5245482](#).
- Ni T, Bojinov I, Zhao J (2023) Design of panel experiments with spatial and temporal interference. [Available at SSRN 4466598](#).
- Pashley NE, Bind MAC (2023) Causal inference for multiple treatments using fractional factorial designs. *Canadian Journal of Statistics* 51(2):444–468.
- Raftery AE (1995) Bayesian model selection in social research. *Sociological methodology* 111–163.
- Rosenman ET, Basse G, Owen AB, Baiocchi M (2023) Combining observational and experimental datasets using shrinkage estimators. *Biometrics* 79(4):2961–2973.
- Simester D, Timoshenko A, Zoumpoulis SI (2025) A sample size calculation for training and certifying targeting policies. *Management Science*.
- Song Y, Sun T (2024) Ensemble experiments to optimize interventions along the customer journey: A reinforcement learning approach. *Management Science* 70(8):5115–5130.

- Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 3, 197–207 (USA: University of California Press).
- Tang D, Agarwal A, O'Brien D, Meyer M (2010) Overlapping experiment infrastructure: More, better, faster experimentation. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 17–26.
- Tetenov A (2016) An economic theory of statistical testing. Technical report, cemmap working paper.
- Ursu RM (2018) The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science* 37(4):530–552.
- Vershynin R (2010) Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*
- Wager S (2024) *Causal inference: A statistical learning approach* (In preparation), URL [https://web.stanford.edu/swager/causal\\_inf\\_book.pdf](https://web.stanford.edu/swager/causal_inf_book.pdf).
- Wainwright MJ (2019) *Basic tail and concentration bounds*, 21–57. Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press), URL <http://dx.doi.org/10.1017/9781108627771.002>.
- Wu CJ, Hamada MS (2011) *Experiments: planning, analysis, and optimization* (John Wiley & Sons).
- Xiong R, Chin A, Taylor S (2023) Bias-variance tradeoffs for designing simultaneous temporal experiments. *The KDD'23 Workshop on Causal Discovery, Prediction and Decision*, 115–131 (PMLR).
- Xiong T, Wang Y, Zheng S (2020) Orthogonal traffic assignment in online overlapping a/b tests.
- Ye Z, Zhang DJ, Zhang H, Zhang R, Chen X, Xu Z (2022) Cold start to improve market thickness on online advertising platforms: Data-driven algorithms and field experiments. *Management Science* 69(7):3838–3860.
- Ye Z, Zhang Z, Zhang D, Zhang H, Zhang RP (2025) Deep-learning-based causal inference for large-scale combinatorial experiments: Theory and empirical evidence. *Management Science*, Forthcoming .
- Zeng Z, Dai H, Zhang DJ, Zhang H, Zhang R, Xu Z, Shen ZJM (2023) The impact of social nudges on user-generated content for social network platforms. *Management Science* 69(9):5189–5208.
- Zhan R, Han S, Hu Y, Jiang Z (2024) Estimating treatment effects under recommender interference: A structured neural networks approach. *arXiv preprint arXiv:2406.14380* .
- Zhang Z, Zeng Z, Zhan R, Zhang D (2025) Personalized policy learning through discrete experimentation: Theory and empirical evidence. Available at SSRN .

## Appendix

### A. Proofs

**Proof of Theorem 1.** First of all, because true ATE  $\tau_k$  follows the normal distribution  $\mathcal{N}(\tau_0, \sigma_0^2)$ ,  $\mathcal{R}(\beta, \tau_0)$  can be written as:

$$\begin{aligned}\mathcal{R}(\beta, \tau_0) &= \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k \in [K]} \tau_k \mathbb{P}\left(\bar{\tau}_k > \frac{N}{N + \beta} \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}}\right) \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\tau_k - \tau_0)^2}{2\sigma_0^2}} \Phi\left(\frac{N\tau_k + \tau_0\beta}{2\sqrt{N}\sigma} - z_{1-\alpha/2}\right) \tau_k d(\tau_k).\end{aligned}$$

We take the derivative of  $\mathcal{R}(\beta, \tau_0)$  with respect to  $\beta$  as follows:

$$\frac{d(\mathcal{R}(\beta, \tau_0))}{d(\beta)} = \frac{\exp\left(-\frac{(N\tau_0 + \beta\tau_0 - 2\sqrt{N}z_{1-\alpha/2}\sigma)^2}{2N(N\sigma_0^2 + 4\sigma^2)}\right) \tau_0 \left(-\beta\tau_0\sigma_0^2 + 2\sigma(\sqrt{N}z_{1-\alpha/2}\sigma_0^2 + 2\tau_0\sigma)\right)}{\sqrt{N}\sqrt{2\pi}\sigma_0 \sqrt{\frac{4}{\sigma_0^2} + \frac{N}{\sigma^2}} \sigma (N\sigma_0^2 + 4\sigma^2)}.$$

We can observe that when  $\beta \geq 0$ , the value of the derivative gradually changes from positive to negative. Thus,  $\mathcal{R}(\beta, \tau_0)$  reaches the maximum value when  $\beta$  takes the following value:

$$\beta^* = \frac{4\sigma^2}{\sigma_0^2} + \frac{2\sqrt{N}z_{1-\alpha/2}\sigma}{\tau_0}. \quad (\text{EC.1})$$

This concludes the proof.

**Proof of Theorem 2.** We have, by definition,  $\hat{\tau}_0 = \frac{1}{K} \sum_k \hat{\tau}_k$  and first prove that  $\hat{\tau}_0 \rightarrow_p \tau_0$ . Specifically, we have,

$$\left| \frac{1}{K} \sum_k \hat{\tau}_k - \tau_0 \right| \leq \left| \frac{1}{K} \sum_k \hat{\tau}_k - \frac{1}{K} \sum_k \tau_k \right| + \left| \frac{1}{K} \sum_k \tau_k - \tau_0 \right|.$$

Since the ATE satisfies,  $\hat{\tau}_k \sim \mathcal{N}(\tau_k, \sigma^2)$ , this implies  $\hat{\tau}_k$  is a sub-Gaussian random variable. Thus, by the concentration theorem of sub-Gaussian random variables (Proposition 2.5 in Wainwright (2019)), for any  $t > 0$ , we have,

$$\mathbb{P}\left(\left| \frac{1}{K} \sum_k \hat{\tau}_k - \frac{1}{K} \sum_k \tau_k \right| \geq t\right) \leq 2 \exp\left(-KNt^2/(8\sigma^2)\right).$$

Consequently, we have,  $\left| \frac{1}{K} \sum_k \hat{\tau}_k - \frac{1}{K} \sum_k \tau_k \right| \rightarrow_p 0$  when  $K \rightarrow \infty$ .

In addition, we have,  $\tau_k \sim \mathcal{N}(\tau_0, \sigma_0^2)$ , by the law of large numbers, we have  $\left| \frac{1}{K} \sum_k \tau_k - \tau_0 \right| \rightarrow_p 0$ . Combining the above results, we can conclude that  $\left| \frac{1}{K} \sum_k \hat{\tau}_k - \tau_0 \right| \rightarrow_p 0$ , and equivalently,  $\hat{\tau}_0 \rightarrow_p \tau_0$ . Then, we can conclude that  $\hat{\tau}_0 \rightarrow_p \tau_0$ .

In order to prove  $\hat{\beta}^* \rightarrow_p \beta^*$ , we first prove the following convergence results:

$$\begin{aligned}\frac{1}{K} \sum_k s_k^2 &\rightarrow_p \sigma^2, \\ \frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 &\rightarrow_p \sigma_0^2 + \frac{4\sigma^2}{N}.\end{aligned}$$

We will first prove the convergence of variance for the treatment group as follows:

$$\frac{1}{K} \sum_k \frac{1}{\frac{N}{2} - 1} \sum_{D_{k,i}=1} \left( Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=1} Y_{k,i} \right)^2 \rightarrow_p \sigma^2. \quad (\text{EC.2})$$

Because the error term  $\epsilon_k$  follow the normal distribution,  $Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=j} Y_{k,i}$  also follows the normal distribution which makes that  $(Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=j} Y_{k,i})^2$  is a sub-exponential random variable according to Lemma 5.14 of Vershynin (2010). Combining the fact that  $\mathbb{E}\left[\frac{1}{\frac{N}{2}-1} \sum_{D_{k,i}=1} (Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=j} Y_{k,i})^2\right] = \sigma^2$  and the independence across all experiments, by applying the concentration theorem of sub-exponential random variables (Proposition 2.9 in Wainwright (2019)), we can prove that Eqn. (EC.2) holds. The analysis in the control group is the same as the above one for the treatment group. Thus, we have,

$$\frac{1}{K} \sum_k s_k^2 = \frac{1}{N-2} \frac{1}{K} \sum_k \sum_{j \in \{0,1\}} \sum_{D_{k,i}=j} \left( Y_{k,i} - \frac{2}{N} \sum_{D_{k,i}=j} Y_{k,i} \right)^2 \xrightarrow{p} \sigma^2. \quad (\text{EC.3})$$

For the term  $\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2$ , we can decompose it as follows:

$$\begin{aligned} \frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 &= \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 + \frac{1}{K} \sum_k (\tau_k - \hat{\tau}_0)^2 + \frac{1}{K} \sum_k 2(\hat{\tau}_k - \tau_k)(\tau_k - \hat{\tau}_0) \\ &= \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 + \frac{1}{K} \sum_k (\tau_k - \tau_0)^2 + \frac{1}{K} \sum_k 2(\tau_k - \tau_0)(\tau_0 - \hat{\tau}_0) \\ &\quad + (\tau_0 - \hat{\tau}_0)^2 + \frac{1}{K} \sum_k 2(\hat{\tau}_k - \tau_k)(\tau_k - \hat{\tau}_0) \\ &= \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 + \frac{1}{K} \sum_k (\tau_k - \tau_0)^2 + \frac{1}{K} \sum_k 2(\tau_k - \tau_0)(\tau_0 - \hat{\tau}_0) \\ &\quad + (\tau_0 - \hat{\tau}_0)^2 + \frac{1}{K} \sum_k 2\tau_k(\hat{\tau}_k - \tau_k) - 2\hat{\tau}_0 \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k). \end{aligned}$$

Since we have  $\mathbb{E}[(\hat{\tau}_k - \tau_k)^2] = \frac{4\sigma^2}{N}$  and  $\mathbb{E}[(\tau_k - \tau_0)^2] = \sigma_0^2$ , by the concentration theorem of sub-exponential random variables (Proposition 2.9 in Wainwright (2019)), we obtain,

$$\begin{aligned} \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 &\xrightarrow{p} \frac{4\sigma^2}{N}, \\ \frac{1}{K} \sum_k (\tau_k - \tau_0)^2 &\xrightarrow{p} \sigma_0^2. \end{aligned}$$

The remaining three parts of the decomposition,  $(\tau_0 - \hat{\tau}_0)^2$ ,  $\frac{1}{K} \sum_k 2\tau_k(\hat{\tau}_k - \tau_k)$ , and  $2\hat{\tau}_0 \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)$  converge to zero in probability due to the fact  $\hat{\tau}_0 \xrightarrow{p} \tau_0$  and  $\mathbb{E}[\hat{\tau}_k - \tau_k] = 0$ . Thus, we can show that,

$$\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 \xrightarrow{p} \sigma_0^2 + \frac{4\sigma^2}{N}. \quad (\text{EC.4})$$

Furthermore, combining the result in Eqn. (EC.3), we can conclude that:

$$\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k 4s_k^2 \xrightarrow{p} \sigma_0^2. \quad (\text{EC.5})$$

Finally, by Slutsky's Theorem, we have,

$$\begin{aligned} \frac{\frac{1}{K} \sum_k 4s_k^2}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{KN} \sum_k 4s_k^2} &\xrightarrow{p} \frac{4\sigma^2}{\sigma_0^2}, \\ \frac{z_{1-\alpha/2} \sqrt{N \frac{1}{K} \sum_k 4s_k^2}}{\hat{\tau}_0} &\xrightarrow{p} \frac{2\sqrt{N} z_{1-\alpha/2} \sigma}{\tau_0}. \end{aligned}$$

Thus, we have,  $\hat{\beta}^* \xrightarrow{p} \beta^*$ . This completes the proof.

**Proof of Theorem 3.** We first prove that  $\bar{r}(\hat{\beta}^*, \hat{\tau}_0) - \tilde{r}(\beta^*, \tau_0) \rightarrow_p 0$  when  $K \rightarrow \infty$ . By definition, we have,

$$\begin{aligned}\bar{r}(\hat{\beta}^*, \hat{\tau}_0) &= \frac{1}{K} \sum_{\hat{\tau}_k > \frac{N}{N+\hat{\beta}^*} \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}}} \tau_k = \frac{1}{K} \sum_k \mathbb{1}\left\{\hat{\tau}_k > \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N}\right\} \tau_k, \\ \tilde{r}(\beta^*, \tau_0) &= \frac{1}{K} \sum_k \mathbb{1}\left\{\hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N}\right\} \tau_k.\end{aligned}$$

Thus, we compute the difference between  $\bar{r}(\hat{\beta}^*, \hat{\tau}_0)$  and  $\tilde{r}(\beta^*, \tau_0)$ ,

$$\left| \bar{r}(\hat{\beta}^*, \hat{\tau}_0) - \tilde{r}(\beta^*, \tau_0) \right| = \left| \frac{1}{K} \sum_k \tau_k \left( \mathbb{1}\left\{\hat{\tau}_k > \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N}\right\} - \mathbb{1}\left\{\hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N}\right\} \right) \right| \quad (\text{EC.6})$$

$$\leq \frac{1}{K} \sum_k |\tau_k| \cdot \left| \mathbb{1}\left\{\hat{\tau}_k > \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N}\right\} - \mathbb{1}\left\{\hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N}\right\} \right| \quad (\text{EC.7})$$

$$= \frac{1}{K} \sum_k |\tau_k| \cdot \mathbb{1}\left\{ \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right\} \quad (\text{EC.8})$$

$$+ \frac{1}{K} \sum_k |\tau_k| \cdot \mathbb{1}\left\{ \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} < \hat{\tau}_k \leq \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} \right\}. \quad (\text{EC.9})$$

We first analyze the indicator variable,  $|\tau_k| \cdot \mathbb{1}\left\{ \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right\}$ . By Markov's inequality, for any  $\xi > 0$ , we have:

$$\mathbb{P}\left(|\tau_k| \cdot \mathbb{1}\left\{ \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right\} \geq \xi\right) \quad (\text{EC.10})$$

$$\leq \frac{1}{\xi} \mathbb{E}_{\tau_k, \hat{\epsilon}_{k,i}} \left[ |\tau_k| \mathbb{1}\left\{ \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right\} \right], \quad (\text{EC.11})$$

where,

$$\begin{aligned}\mathbb{E}_{\tau_k, \hat{\epsilon}_{k,i}} \left[ |\tau_k| \cdot \mathbb{1}\left\{ \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right\} \right] \\ = \mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \mathbb{P}\left( \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \right].\end{aligned}$$

We can find a function,  $\tau_k^2 + 1$ , such that, for any  $K$  and  $\tau_k$ ,

$$|\tau_k| \cdot \mathbb{P}\left( \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \leq |\tau_k| \leq \tau_k^2 + 1. \quad (\text{EC.12})$$

By Assumption 1,  $\tau_k^2 + 1$  is integrable. Thus, by the Dominated Convergence Theorem, we have

$$\begin{aligned}\lim_{K \rightarrow \infty} \mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \mathbb{P}\left( \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \right] \\ = \mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \lim_{K \rightarrow \infty} \mathbb{P}\left( \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \right].\end{aligned}$$

Furthermore, based on the proof and results of Theorem 2, we have,  $\hat{\sigma} \rightarrow_p \sigma$  and  $\hat{\beta}^* \rightarrow_p \beta^*$ . Thus, when  $K \rightarrow \infty$ , we have,

$$\mathbb{P}\left( \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \rightarrow 0. \quad (\text{EC.13})$$

Therefore, we have,

$$\mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \mathbb{P} \left( \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \right] \rightarrow 0. \quad (\text{EC.14})$$

Thus, combining Eqn. (EC.10) and (EC.14), we have,

$$|\tau_k| * 1\{ \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \} \rightarrow_p 0. \quad (\text{EC.15})$$

The analysis for  $|\tau_k| \cdot 1\{ \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} < \hat{\tau}_k \leq \frac{2\hat{\sigma}z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\hat{\tau}_0}{N} \}$  is the same. In the Eqn. (EC.8) and (EC.9), we take the average of the terms associated with indicator variables, and then, we can conclude that  $|\bar{r}(\hat{\beta}^*, \hat{\tau}_0) - \tilde{r}(\beta^*, \tau_0)| \rightarrow_p 0$ . Thus, the proof of  $\bar{r}(\hat{\beta}^*, \hat{\tau}_0) - \tilde{r}(\beta^*, \tau_0) \rightarrow_p 0$  is completed.

We next prove the convergence result,  $\tilde{r}(\beta^*, \tau_0) \rightarrow_p \mathcal{R}(\beta^*, \tau_0)$ . By definition, we have,

$$\mathcal{R}(\beta^*, \tau_0) = \lim_{K \rightarrow \infty} \mathbb{E}(\tilde{r}(\beta^*, \tau_0)) = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \tau_k. \quad (\text{EC.16})$$

Subsequently, we have the following:

$$\begin{aligned} & \frac{1}{K} \sum_k 1\left\{ \hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right\} \tau_k - \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \tau_k \rightarrow_p 0 \\ & \Rightarrow \tilde{r}(\beta^*, \tau_0) - \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \tau_k \rightarrow_p 0. \end{aligned}$$

The above relationship holds because of the fact that  $1\{ \hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \}$  is a sub-Gaussian random variable and the concentration theorem of sub-Gaussian random variables (Proposition 2.5 in Wainwright (2019)). Thus, combining the result  $\bar{r}(\hat{\beta}^*, \hat{\tau}_0) - \tilde{r}(\beta^*, \tau_0) \rightarrow_p 0$ , we can conclude that:

$$\bar{r}(\hat{\beta}^*, \hat{\tau}_0) - \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N} \right) \tau_k \rightarrow_p 0. \quad (\text{EC.17})$$

In the proof of Theorem 1, we show that  $\mathcal{R}(\beta^*, \tau_0)$  is a constant. By the definition of  $\mathcal{R}(\beta^*, \tau_0)$ , we have,  $\frac{1}{K} \sum_k \mathbb{P}(\hat{\tau}_k > \frac{2\sigma z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*\tau_0}{N}) \tau_k \rightarrow_p \mathcal{R}(\beta^*, \tau_0)$ . Thus, combining the result in Eqn. (EC.17), we have,  $\bar{r}(\hat{\beta}^*, \hat{\tau}_0) \rightarrow_p \mathcal{R}(\beta^*, \tau_0)$ . This completes the proof.

**Proof of Theorem 4.** First, the true ATE  $\tau_k$  follows the normal distribution  $\mathcal{N}(\tau_0, \sigma_0^2)$ , and  $b_k$  follows the same distribution. Second,  $b_k$  and  $\tau_k$  are independent. Thus,  $\mathcal{R}(\beta, \tau_0)$  can be computed as:

$$\begin{aligned} \mathcal{R}(\beta, \tau_0) &= \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k \in [K]} \tau_k \mathbb{P} \left( \bar{\tau}_k > \frac{N}{N + \beta} \frac{\sigma b_k}{\sqrt{N}} z_{1-\alpha/2} \right) \\ &= \mathbb{E}_{b_k} \left[ \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\tau_k - \tau_0)^2}{2\sigma_0^2}} \Phi \left( \frac{N\tau_k + \tau_0\beta}{\sqrt{N}\sigma b_k} - z_{1-\alpha/2} \right) \tau_k d(\tau_k) \right]. \end{aligned}$$

Conditioning on  $b_k$ , we take the derivative of the realized  $\mathcal{R}(\beta, \tau_0)$  with respect to  $\beta(\cdot)$  as follows:

$$\frac{\exp \left( -\frac{(N\tau_0 + \beta\tau_0 - \sqrt{N}z_{1-\alpha/2}\sigma b_k)^2}{2N(N\sigma_0^2 + \sigma^2 b_k^2)} \right) \tau_0 \left( -\beta\tau_0\sigma_0^2 + \sigma b_k \left( \sqrt{N}z_{1-\alpha/2}\sigma_0^2 + \tau_0\sigma b_k \right) \right)}{b_k \sqrt{N} \sqrt{2\pi} \sigma_0 \sqrt{\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2 b_k^2}} \sigma (N\sigma_0^2 + b_k^2\sigma^2)}.$$

We can observe that when  $\beta(b_k) \geq 0$  and increases, the value of the derivative gradually changes from positive to negative. Thus,  $\mathcal{R}(\beta(\cdot), \tau_0)$  reaches the maximum value when  $\beta(\cdot)$  takes the following value:

$$\beta^*(b_k) = \frac{\sigma^2 b_k^2}{\sigma_0^2} + \frac{\sqrt{N}z_{1-\alpha}\sigma b_k}{\tau_0}. \quad (\text{EC.18})$$

This concludes the proof.

**Proof of Theorem 5.** We first prove that  $\hat{\beta}^*(b_k) \rightarrow_p \beta^*(b_k)$ . We first prove that  $\hat{\tau}_0 = \frac{1}{K} \sum_k \hat{\tau}_k \rightarrow_p \tau_0$  as follows:

$$\left| \frac{1}{K} \sum_k \hat{\tau}_k - \tau_0 \right| \leq \left| \frac{1}{K} \sum_k \hat{\tau}_k - \frac{1}{K} \sum_k \tau_k \right| + \left| \frac{1}{K} \sum_k \tau_k - \tau_0 \right|.$$

Since  $\hat{\tau}_k \sim \mathcal{N}(\tau_k, \frac{\sigma^2 b_k^2}{N})$ , we have,  $\hat{\tau}_k$  is a sub-Gaussian random variable. Thus, by the concentration theorem of sub-Gaussian random variables (Proposition 2.5 in Wainwright (2019)), for any  $t > 0$ , we have,

$$\mathbb{P}\left(\left| \frac{1}{K} \sum_k \hat{\tau}_k - \frac{1}{K} \sum_k \tau_k \right| \geq t\right) \leq 2 \exp(-KNt^2/(8\sigma^2)).$$

We then deduce that,  $\left| \frac{1}{K} \sum_k \hat{\tau}_k - \frac{1}{K} \sum_k \tau_k \right| \rightarrow_p 0$  when  $K \rightarrow \infty$ .

Since  $\tau_k \sim \mathcal{N}(\tau_0, \sigma_0^2)$ , by the law of large numbers, we have,  $\left| \frac{1}{K} \sum_k \tau_k - \tau_0 \right| \rightarrow_p 0$ . Combining the above results, we can conclude that  $\left| \frac{1}{K} \sum_k \hat{\tau}_k - \tau_0 \right| \rightarrow_p 0$ , which proves  $\hat{\tau}_0 \rightarrow_p \tau_0$ . This concludes the proof of  $\hat{\tau}_0 \rightarrow_p \tau_0$ .

In order to prove,  $\hat{\beta}^*(b_k) \rightarrow_p \beta^*(b_k)$  given  $b_k$ , we first prove the following convergence results:

$$\frac{1}{K} \sum_k s_k^2 \rightarrow_p \sigma^2, \tag{EC.19}$$

$$\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 \rightarrow_p \sigma_0^2 + \frac{\frac{1}{K} \sum_k b_k^2 \sigma^2}{N}. \tag{EC.20}$$

In addition,  $s_k^2$  is proportional to the chi-squared distribution (see, Page 14 in Amemiya 1985),

$$s_k^2 \sim \frac{\sigma^2}{N - d_x - 1} \cdot \chi_{N-d_x-1}^2. \tag{EC.21}$$

Thus, we have  $\mathbb{E}[s_k^2] = \sigma^2$  and  $s_k^2$  follows sub-exponential distribution (see, Example 2.4 in Wainwright 2019).

By applying the concentration theorem of sub-exponential random variables (Proposition 2.9 in Wainwright (2019)), we have,  $\frac{1}{K} \sum_k s_k^2 \rightarrow_p \sigma^2$ .

For the term  $\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2$ , we can decompose it as follows:

$$\begin{aligned} \frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 &= \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 + \frac{1}{K} \sum_k (\tau_k - \hat{\tau}_0)^2 + \frac{1}{K} \sum_k 2(\hat{\tau}_k - \tau_k)(\tau_k - \hat{\tau}_0) \\ &= \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 + \frac{1}{K} \sum_k (\tau_k - \tau_0)^2 + \frac{1}{K} \sum_k 2(\tau_k - \tau_0)(\tau_0 - \hat{\tau}_0) \\ &\quad + (\tau_0 - \hat{\tau}_0)^2 + \frac{1}{K} \sum_k 2(\hat{\tau}_k - \tau_k)(\tau_k - \hat{\tau}_0) \\ &= \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 + \frac{1}{K} \sum_k (\tau_k - \tau_0)^2 + \frac{1}{K} \sum_k 2(\tau_k - \tau_0)(\tau_0 - \hat{\tau}_0) \\ &\quad + (\tau_0 - \hat{\tau}_0)^2 + \frac{1}{K} \sum_k 2\tau_k(\hat{\tau}_k - \tau_k) - 2\hat{\tau}_0 \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k). \end{aligned}$$

We know that  $\mathbb{E}[(\hat{\tau}_k - \tau_k)^2] = \frac{b_k^2 \sigma^2}{N}$  and  $\mathbb{E}[(\tau_k - \tau_0)^2] = \sigma_0^2$ . Thus, by the concentration theorem of sub-exponential random variables (Proposition 2.9 in Wainwright (2019)), we have,

$$\begin{aligned} \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)^2 &\rightarrow_p \frac{\frac{1}{K} \sum_k b_k^2 \sigma^2}{N}, \\ \frac{1}{K} \sum_k (\tau_k - \tau_0)^2 &\rightarrow_p \sigma_0^2. \end{aligned}$$

The remaining three parts in the decomposition,  $(\tau_0 - \hat{\tau}_0)^2$ ,  $\frac{1}{K} \sum_k 2\tau_k(\hat{\tau}_k - \tau_k)$ , and  $2\hat{\tau}_0 \frac{1}{K} \sum_k (\hat{\tau}_k - \tau_k)$  will converge to zero in probability because of the fact that  $\hat{\tau}_0 \rightarrow_p \tau_0$  and  $\mathbb{E}[\hat{\tau}_k - \tau_k] = 0$ . Thus, we have,

$$\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 \rightarrow_p \sigma_0^2 + \frac{\frac{1}{K} \sum_k b_k^2 \sigma^2}{N}.$$

Furthermore, combining the result in Eqn. (EC.19), we can conclude that:

$$\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{N} \left( \frac{1}{K} \sum_k s_k^2 * \frac{1}{K} \sum_k b_k^2 \right) \rightarrow_p \sigma_0^2.$$

Thus, by Slutsky's Theorem, given  $b_k$ , we can conclude that:

$$\begin{aligned} & \frac{b_k^2 \frac{1}{K} \sum_k s_k^2}{\frac{1}{K} \sum_k (\hat{\tau}_k - \hat{\tau}_0)^2 - \frac{1}{N} \left( \frac{1}{K} \sum_k s_k^2 * \frac{1}{K} \sum_k b_k^2 \right)} \rightarrow_p \frac{\sigma^2 b_k^2}{\sigma_0^2}, \\ & \frac{z_{1-\alpha/2} b_k \sqrt{N \frac{1}{K} \sum_k s_k^2}}{\hat{\tau}_0} \rightarrow_p \frac{\sqrt{N} b_k z_{1-\alpha/2} \sigma}{\tau_0}. \end{aligned}$$

Thus, given  $b_k$ , the proof of  $\hat{\beta}^*(b_k) \rightarrow_p \beta^*(b_k)$  is completed.

We next prove  $\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) \rightarrow_p \mathcal{R}(\beta^*(\cdot), \tau_0)$ . Given  $\{b_1, \dots, b_K\}$ , we first prove that  $\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) - \tilde{r}(\beta^*(\cdot), \tau_0) \rightarrow_p 0$  when  $K \rightarrow \infty$ . By definition, we have,

$$\begin{aligned} \bar{r}(\hat{\beta}^*(b_k), \hat{\tau}_0) &= \frac{1}{K} \sum_{\substack{\hat{\tau}_k > \frac{N}{N+\hat{\beta}^*(b_k)} \cdot \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}}} \tau_k = \frac{1}{K} \sum_k \mathbb{1} \left\{ \hat{\tau}_k > \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} \right\} \tau_k, \\ \tilde{r}(\beta^*(b_k), \tau_0) &= \frac{1}{K} \sum_k \mathbb{1} \left\{ \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \tau_k. \end{aligned}$$

We then compute the difference between  $\bar{r}(\hat{\beta}^*(b_k), \hat{\tau}_0)$  and  $\tilde{r}(\beta^*(b_k), \tau_0)$ ,

$$\begin{aligned} & \left| \bar{r}(\hat{\beta}^*(b_k), \hat{\tau}_0) - \tilde{r}(b_k), \tau_0) \right| \\ &= \left| \frac{1}{K} \sum_k \tau_k \left( \mathbb{1} \left\{ \hat{\tau}_k > \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} \right\} - \mathbb{1} \left\{ \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \right) \right| \\ &\leq \frac{1}{K} \sum_k |\tau_k| \cdot \left| \mathbb{1} \left\{ \hat{\tau}_k > \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} \right\} - \mathbb{1} \left\{ \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \right| \\ &= \frac{1}{K} \sum_k |\tau_k| \cdot \mathbb{1} \left\{ \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \quad (\text{EC.22}) \end{aligned}$$

$$+ \frac{1}{K} \sum_k |\tau_k| \cdot \mathbb{1} \left\{ \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} < \hat{\tau}_k \leq \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} \right\}. \quad (\text{EC.23})$$

We first analyze the term,  $|\tau_k| \cdot \mathbb{1} \left\{ \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\}$ . By Markov's inequality, for any  $\xi > 0$ , we have:

$$\begin{aligned} & \mathbb{P} \left( |\tau_k| \cdot \mathbb{1} \left\{ \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \geq \xi \right) \\ &\leq \frac{1}{\xi} \mathbb{E}_{\tau_k, \hat{\epsilon}_k, i} \left[ |\tau_k| \cdot \mathbb{1} \left\{ \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \right], \quad (\text{EC.24}) \end{aligned}$$

where

$$\begin{aligned} & \mathbb{E}_{\tau_k, \hat{\epsilon}_k, i} \left[ |\tau_k| \cdot \mathbb{1} \left\{ \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \right] \\ &= \mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \mathbb{P} \left( \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \right]. \end{aligned}$$

There exists a function  $\tau_k^2 + 1$ , for any  $K$  and  $\tau_k$ , such that

$$|\tau_k| \cdot \mathbb{P} \left( \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \leq |\tau_k| \leq \tau_k^2 + 1. \quad (\text{EC.25})$$

By Assumption 1,  $\tau_k^2 + 1$  is integrable. Thus, by the Dominated Convergence Theorem, we have

$$\begin{aligned} & \lim_{K \rightarrow \infty} \mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \mathbb{P} \left( \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \right] \\ &= \mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \lim_{K \rightarrow \infty} \mathbb{P} \left( \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \right]. \end{aligned}$$

Furthermore, based on the proof and results of Theorem 5, we have,  $\hat{\sigma} \rightarrow_p \sigma$  and  $\hat{\beta}^*(b_k) \rightarrow_p \beta^*(b_k)$ . Thus, when  $K \rightarrow \infty$ , we have,

$$\mathbb{P} \left( \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \rightarrow 0, \quad (\text{EC.26})$$

and

$$\mathbb{E}_{\tau_k} \left[ |\tau_k| \cdot \mathbb{P} \left( \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \right] \rightarrow 0. \quad (\text{EC.27})$$

Thus, according to Eqn. (EC.24) and (EC.27), we have,

$$|\tau_k| \cdot \mathbb{1} \left\{ \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} < \hat{\tau}_k \leq \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \rightarrow_p 0. \quad (\text{EC.28})$$

The analysis for the term in Equation (EC.23),  $|\tau_k| \cdot \mathbb{1} \left\{ \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} < \hat{\tau}_k \leq \frac{\hat{\sigma} b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\hat{\beta}^*(b_k) \hat{\tau}_0}{N} \right\}$ , is the same. Following Eqn. (EC.22) and (EC.23), we take the average of terms associated with indicator variables, and then, we have,  $|\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) - \bar{r}(\beta^*(\cdot), \tau_0)| \rightarrow_p 0$ . Thus, the proof of  $\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) - \bar{r}(\beta^*(\cdot), \tau_0) \rightarrow_p 0$  is completed.

We next prove  $\tilde{r}(\beta^*(\cdot), \tau_0) \rightarrow_p \mathcal{R}(\beta^*(\cdot), \tau_0)$ . By definition, we have,

$$\mathcal{R}(\beta^*(\cdot), \tau_0) = \lim_{K \rightarrow \infty} \mathbb{E}(\tilde{r}(\beta^*(\cdot), \tau_0)) = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \tau_k. \quad (\text{EC.29})$$

Thus, we have,

$$\begin{aligned} & \frac{1}{K} \sum_k \mathbb{1} \left\{ \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\} \tau_k - \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \tau_k \rightarrow_p 0 \\ & \Rightarrow \tilde{r}(\beta^*(\cdot), \tau_0) - \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \tau_k \rightarrow_p 0. \end{aligned}$$

The above relationship holds because of the fact that  $\mathbb{1} \left\{ \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right\}$  is a sub-Gaussian random variable and the concentration theorem of sub-Gaussian random variables (Proposition 2.5 in Wainwright (2019)). Thus, combining the result  $\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) - \bar{r}(\beta^*(\cdot), \tau_0) \rightarrow_p 0$ , we have,

$$\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) - \frac{1}{K} \sum_k \mathbb{P} \left( \hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N} \right) \tau_k \rightarrow_p 0. \quad (\text{EC.30})$$

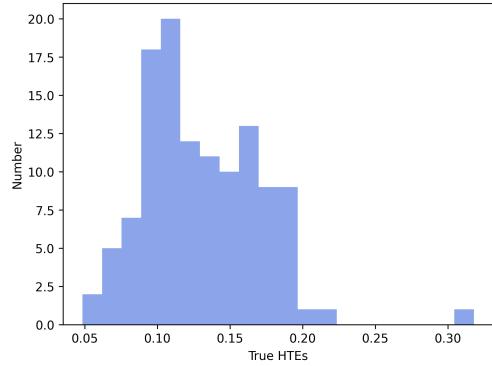
In the proof of Theorem 4, we show that  $\mathcal{R}(\beta^*(\cdot), \tau_0)$  is a constant. By the definition of  $\mathcal{R}(\beta^*(\cdot), \tau_0)$ , we have,  $\frac{1}{K} \sum_k \mathbb{P}(\hat{\tau}_k > \frac{\sigma b_k z_{1-\alpha/2}}{\sqrt{N}} - \frac{\beta^*(b_k) \tau_0}{N}) \tau_k \rightarrow_p \mathcal{R}(\beta^*(\cdot), \tau_0)$ . Thus, combining the result in Eqn. (EC.30), we can deduce that  $\bar{r}(\hat{\beta}^*(\cdot), \hat{\tau}_0) \rightarrow_p \mathcal{R}(\beta^*(\cdot), \tau_0)$  and this completes the proof.

## B. Numerical Results with Expedia Experiment Data

In this dataset, we complement our analysis with an additional dataset to assess the performance of the DPTR method in a non-overlapping setting. The dataset comes from a field experiment conducted by Expedia, the world's largest online travel agency. This publicly available dataset<sup>6</sup> includes data from consumers searching for hotels, who were randomly assigned to one of two groups: (i) those who viewed a personalized ranking, where hotels were ordered according to their suitability for consumers based on Expedia's internal ranking algorithm, and (ii) those who viewed a random ranking, where hotels were listed in no particular order.

Following the data cleaning process outlined in Ursu (2018), we obtained a total of approximately 166 thousands of consumer queries for hotels, along with their corresponding choices (clicks and purchases), spanning an eight-month period ending in June 2013. This dataset includes approximately 4.5 million observations of hotels displayed on Expedia.

In this dataset, we use the origin-destination pair (the country of the customer and the country of the hotels being searched) to partition the whole dataset. To ensure each group has enough data for reliable analysis, we exclude groups with fewer than 250 queries. This results in 119 groups, that is,  $K = 119$ . The histogram of the “true” HTE across all groups, estimated from the entire dataset, is shown in Figure EC.1.



**Figure EC.1    Histogram of true HTEs for all groups of the Expedia dataset.**

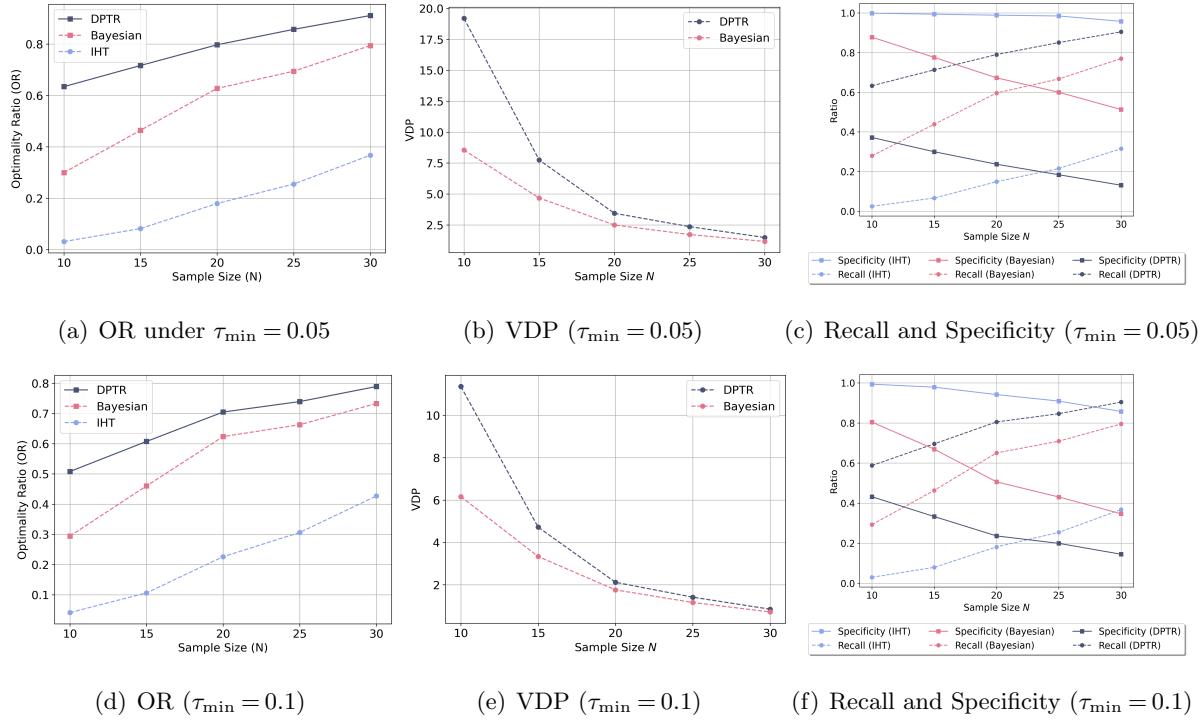
Considering that in real-world scenarios, platforms incur costs  $\tau_{\min}$  for both the design and maintenance of algorithms. Thus, for any policy  $k$ , the actual reward obtained by the platform is given by,  $(\tau_k - \tau_{\min})$ . We assess the impact of implementation costs in this example by considering two frictional cost values,  $\tau_{\min} = 0.05$  and  $0.1$ <sup>7</sup>. Since the data sizes vary across groups, we normalize  $(\tau_k - \tau_{\min})$  when calculating the OR and VDP values by multiplying it by the parameter  $\frac{N_{k,0} + N_{k,1}}{\sum_k (N_{k,0} + N_{k,1})}$ , which represents the proportion of group  $k$ 's data size relative to the total data size of all groups.

Besides the OR and VDP metrics, we also report Recall rate which measures the ratio of correct decisions made by the roll-out method (IHT or DPTR) across all experiments where the roll-out should occur and

<sup>6</sup> [www.kaggle.com/c/expedia-personalized-sort/data](http://www.kaggle.com/c/expedia-personalized-sort/data)

<sup>7</sup> This does not imply that the actual implementation cost of recommendation algorithms reaches this level, although sophisticated recommendations may increase webpage response times, potentially leading to long-term negative outcomes for the platform, as webpage speed is an important factor for online consumers (Gallino et al. 2023).

Specificity which measures the ratio of correct decisions made by the roll-out method (**IHT** or **DPTR**) across all experiments where the roll-out should not occur.



**Figure EC.2 Performance comparisons under different sample size  $N$  and  $\tau_{\min}$  in Expedia dataset.**

Furthermore, for any  $\tau_{\min}$ , we vary the sample size  $N$  from 10 to 30 in increments of 5 and repeat the experiment 1,000 times. The averaged results are shown in Figure EC.2. As depicted in the figure, we first observe that, regardless of the value of  $\tau_{\min}$ , the trends across all indicators remain largely consistent, with no significant fluctuations. This suggests that the choice of  $\tau_{\min}$  has minimal impact on the comparison between the **IHT** and **DPTR** methods.

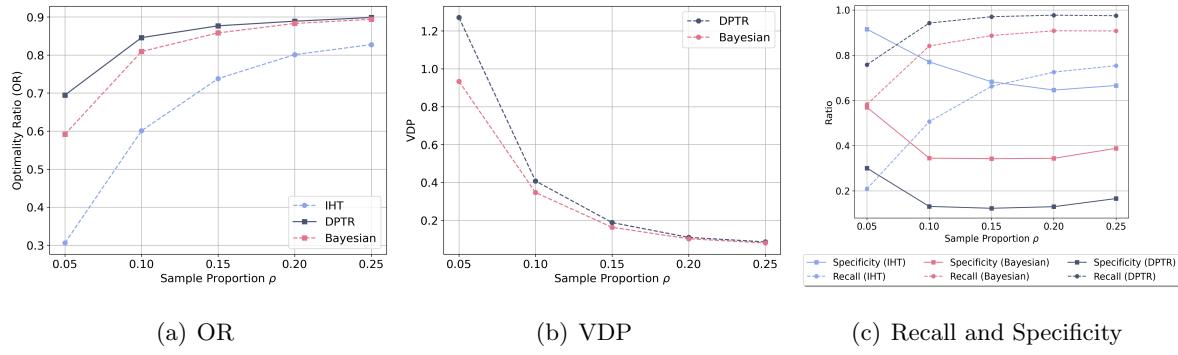
According to Figure EC.2(a) and (d), the **DPTR** method consistently outperforms the **IHT** and **Bayesian** methods, irrespective of the sample size  $N$ . Additionally, as the sample size increases, the reduction in estimator variance leads to a corresponding increase in OR values for both methods. Figure EC.2(b) and (e) further reveal that the relative performance advantage of the **DPTR** and **Bayesian** methods over the **IHT** method increases significantly as the amount of experimental data decreases. This finding highlights that, when experimental data is limited, the data-pooling methods, such as **DPTR** and **Bayesian** methods, can provide substantial performance improvements.

Finally, Figure EC.2(c) and (f) show that when  $N$  is small, the Recall value under the **IHT** method is closer to 0, indicating that few or no personalized recommendations are rolled out to specific groups. Since few or no groups launch the personalized recommendation, the Specificity value is higher under the **IHT** method. On the other hand, despite large variation in the estimate of the benefit of personalized recommendations in each group, the pooled estimate in the **DPTR** and **Bayesian** methods reduce this variation, pushing the

estimate towards a more positive region. As a result, the Recall value under the DPTR and Bayesian methods increases significantly. This comes at the cost of some groups being incorrectly selected for the personalized recommendation. However, overall, by balancing Recall and Specificity, the DPTR and Bayesian methods generate a much higher reward compared to the IHT method in this tailored roll-out decision situation.

We observe that the IHT method has a very low Recall value and a high Specificity value. This phenomenon suggests that traditional methods (IHT) tend to exclude experiments that should not be rolled out, even if this tendency results in the exclusion of many experiments that should actually be rolled out. In contrast, our proposed DPTR method maximizes the reward from the final roll-out decision by balancing the trade-off between Recall and Specificity.

Furthermore, we test the performance of our method when different groups have different sample sizes. We denote  $\rho$  as the sample proportion with respect to the total data size, which means we will randomly select  $\rho \times (N_{k,1} + N_{k,0})$  customer queries as sample size for group  $k$ . We vary  $\rho$  from 0.05 to 0.25 in increments of 0.05 while keeping  $\tau_{\min} = 0.1$  fixed. Similarly, we repeat the experiment 1,000 times and present the averaged results in Figure EC.3. We observe that variations in sample sizes across different groups do not affect the performance of our DPTR method.



**Figure EC.3 Performance comparisons under different sample proportion  $\rho$  in Expedia dataset.**